Eureka Digital Archive

archim.org.uk/eureka

This work is published under the CC BY 4.0 license.

https://creativecommons.org/licenses/by/4.0/

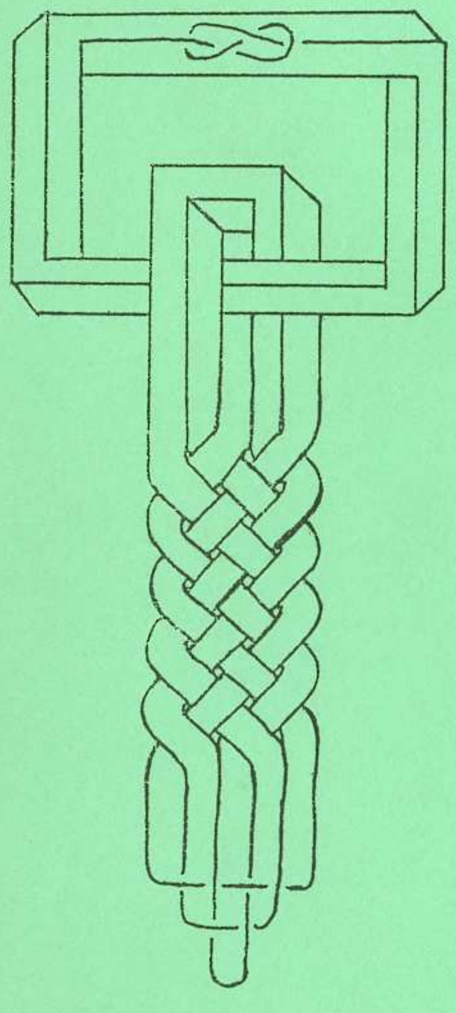
Eureka Editor
The Archimedeans
Centre for Mathematical Sciences
Wilberforce Road
Cambridge CB3 0WA
United Kingdom

archim-eureka@srcf.net

Published by The Archimedeans, the mathematics student society of the University of Cambridge

Thanks to the Betty & Gordon Moore Library, Cambridge

EUREKA 48



MARCH 1988

EUREKA 48

Eureka is the journal of the Archimedeans, which is the Cambridge University Mathematical Society, and a Junior Branch of the Mathematical Association. Eureka is published approximately annually, but since it, like the Archimedeans, is run entirely by student volunteers, it is impossible to guarantee precise publication dates. The society also publishes QARCH, a problems journal.

Archimedeans Committee 1987-8

President	Graham Nelson	Selwyn
Vice-President	Matthew Richards	Trinity
Secretary	Alan Stacey	St. Catharine's
Junior Treasurer	Paul Glover	Churchill
Registrar	Ian Redfern	Trinity
Publicity Manager	Thomas Bending	Trinity
Chronicler	Christopher Fewster	Churchill
Entertainments Manager	Simon Morris	Trinity
Senior Treasurer	Dr. J.R. Partington	Fitzwilliam

Subscriptions

A subscription account may be opened by sending at least five pounds to the Business Manager at the address below; a copy of each issue will be sent to you when it is published and you will be informed when your account runs low. It is also possible to subscribe to Eureka by by standing order. Back numbers of certain issues are available. Photocopies of any issue may be supplied by special arrangement. Write to the Business Manager for details.

Electronic Mail

The Society can receive electronic mail via JANET, ARPAnet, etc, and can read most electronic media. Our electronic address is given below. There may be long delays in dealing with messages (electronic or otherwise) during University vacations (especially June to October), and you are asked not to send large amounts of text electronically without prior warning.

Eureka

Editor	Simon Morris	Trinity
Business Manager	Paul Glover	Churchill

Correspondence, clearly marked "Editor" or "Business Manager"

The Arts School
Bene't Street
Cambridge CB2 3PY
ENGLAND

Electronic mail

archim%camphx@caga (UK)
archim%camphx%caga@ucl-cs.arpa (USA)

CONTENTS

Editorial 1

Truth in Mathematics? (Simon Morris) 2

The Society (Christopher J. Fewster) 7

The Eccentricity of Mathematicians (John F. Bowers) 8

Colouring Patterns of Squares (Graham Nelson) 15

Wit and Wisdom I 19

The Power of Mathematics (Lynn Arthur Steen) 20

Mathematics Has A Front And A Back (Reuben Hersh) 27

Backgammon (Stuart Melbourne) 32

The Propositional Calculus With Directed Graphs (Martin Gardner and Frank Harary) 34

Wit and Wisdom II 41

A General Survey of the Mathematical Philosophy of Immanuel Kant (Robert K. Clifton) 42

Problems Drive (Chris Kilgour and Matthew Richards) 55

Introduction to Doughalese (Douglas R. Hofstadter) 57

An Introduction to Wargaming (Paul Fernandez) 65

Academic Freedom - A Personal View of Life Beyond the PhD (John Skilling) 68

Differentiation Rules, Okay (Clive Monk) 71

Ramanujan - A Glimpse of His Life and His Mathematics (Béla Bollobás) 81

Answers to Problems Drive 99

EDITORIAL

Welcome to Eureka 48 - the fullest Eureka ever published. A glance at the contents page will reveal the diversity of topics included; mathematics itself, the philosophy of mathematics, the history of mathematics, the way mathematics is practised - there are even articles with no mathematics at all!

This appeal to as wide a range of interests as possible is the essence of Eureka. Some 70% of undergraduate mathematics students at Cambridge are Archimedeans, yet only a small core of this number regularly attends Archimedean events, and many will only go to a few in their undergraduate career. Eureka is for all Archimedeans, not just the devoted few.

I would like to thank Thomas Bending for the front cover and helping with the typing, all the contributors for their time and effort, and the Archimedeans for letting me edit their journal.

Simon Morris, March 1988

TRUTH IN MATHEMATICS ?

Simon Morris (Trinity College, Cambridge)

INTRODUCTION

The three main philosophies of mathematics[1] are Platonism[2,3], formalism[4,5], and constructivism[6,7] (or intuitionism). I will not discuss constructivism in detail, much though it deserves it.

Platonism proposes that mathematical statements describe a realm of mind-independent abstract objects. This raises the question of how mathematicians get access to this realm - is there some mathematically sensitive region of the brain : the pineal gland, for example? Other difficulties are the relevance of proof, and the nature of mistakes in Platonism.

Formalism proposes that a mathematical system is composed of a set of axioms and a set of rules of inference. A proof is an ordered set of statements such that every element of that set either is an axiom, or may be derived from one or more of the preceding elements using one of the rules of inference. A true statement in the system is a statement such that there exists a proof with that statement as its last element. (I have skipped the details of "rules of inference" and "derivation" - these can be made entirely rigorous[8]).

This statement of formalism makes it clear that formalism is a mathematician's philosophy of mathematics, so it is perhaps surprising that it accords so poorly with the way mathematicians do mathematics. Axioms are unquestionably important, and indeed *de rigeur* for mathematics[9], but only mathematical logic is ever concerned with so much as stating rules of inference, and to explicitly refer to one of a given set of rules of inference at every point of a proof would be impossibly tedious. Certainly formalism has proved indispensable for metamathematics. This is, roughly speaking, the mathematical study of mathematics, and includes results such as Godel's incompleteness theorems[10], and the Godel-Cohen theorem on the undecidability of the continuum hypothesis[11,12]. However, in this context, formalism is really a model of mathematics (I use "model", oddly enough, almost in the sense in which it is used in applied mathematics). I conclude that formalism is inadequate as a philosophical (as opposed to a merely mathematical) description of mathematics.

The object of the philosophy of mathematics is to clarify the nature of mathematics; but historically, most effort has been spent on the attempt to clarify the nature of mathematical truth. This points towards two prevalent assumptions; firstly, that mathematics is an enterprise geared towards the production of truth, and secondly, that this truth is fundamental, eternal, and certain. Encouraged by Lakatos' philosophy of dubitability[13,14,15], a reaction against the latter assumption has become very visible, but the former is harder to dislodge. However, the two assumptions are linked by the notion of "truth", and therein lies the central dilemma of the modern philosophy of mathematics. The impregnability of mathematical truth is crumbling in every direction; formalism is inadequate, proofs by computer remain controversial[16,17,18,19,20], there are contradictory theorems, each with an apparently flawless proof[19], and I have even heard an eminent mathematician say that, although unable to locate an error, or any particular weak point, in a certain lengthy proof, he was not convinced of the proof's correctness! But if mathematical truth is uncertain, *what is mathematics good for?*

THESIS

I will avoid both horns of the dilemma by avoiding both assumptions and proposing this truth-free definition of mathematics :

Mathematics is the study of precise ideas.

JUSTIFICATION

It is immediately clear that this makes some sort of sense. A good way to study precise ideas is to study the relations between precise ideas. One might reasonably start with a small core of precise ideas and see which other ideas are related (in some necessarily precise way) to these core ideas. This is essentially the formalistic method. Of course, it is only one way of studying precise ideas, and this definition of mathematics may seem excessively broad. However, this century has seen the birth of constructivism, a completely novel form of mathematics. We have no guarantee that mathematics will not continue to be broadened in unexpected directions.

I will also justify my definition by proposing similar definitions for other fields of endeavour. Philosophy is the study of ideas. Science is the abstraction of precise ideas from the world. Applied mathematics is the study of these particular ideas (remember that both science and mathematics have, at some time, been regarded as part of philosophy). Philosophical ideas, however, have the additional characteristic of referring to some object external and antecedent to philosophy itself. This is a property not shared by mathematical ideas. The extent to which the philosophical object either is precise, or can have useful, precise ideas attached to it, is invariably a matter for debate. This is a problem with which mathematicians do not have to contend.

I will try to clarify these definitions with an example. Consider the statement "The fence is green". The philosopher will investigate the meaning of the statement, the meaning of meaning, the nature of "fence" and of "green", and so on. The scientist will, after appropriate measurement, say that if light containing sufficiently high frequencies is incident on the fence, light of certain frequencies will be reflected from the fence. He will also be able to explain these terms very precisely. What, then, is so imprecise about the scientist's idea that it should be deemed non-mathematical? The crucial element of the idea is the fence. Whether considered as a thing-in-itself or as another idea of the scientist's, it is imprecise; external and antecedent to her precise thoughts on colour and its attachment to the fence. She cannot say precisely what she means by "fence"; she can only point. Precise ideas, then, can only refer to precise ideas - mathematics is isolated. Having concluded that scientific ideas are imprecise, I must expand on my assertion that scientists abstract precise ideas from the world. Scientists' ideas come in pairs. Consider, for example, the idea that light is waves in the electromagnetic field. The precise, hence abstract component of the idea is the waveform solution of Maxwell's equations in a vacuum - an essentially mathematical idea. The imprecise component is the idea that this solution characterises light.

PLATONISM AND FORMALISM

It is possible to place Platonism and formalism on a spectrum of the precise idea; Platonism emphasises the idea (at the expense of any but a geographical characterisation of the idea), whereas formalism emphasises precision (at the expense of the idea). Consider :

The essence of Platonism is the reality, one might almost say the solidity, of mathematical objects. However, it places them in some unexplained metaphysical realm, which is unsatisfactory. I find no alternative but to locate them in the psyche, where they adopt the status of ideas. To explain the (at least relative) objectivity of mathematical ideas, some characterisation of these ideas is needed. I suggest precision. Davis and Hersh[21] have proposed an identical philosophy, up to this point, but their characteristic is reproducibility. "The fence is green" is a reproducible idea, but it is not mathematical. We can even, with the help of a scientist, reproduce this idea in the mind of a small, furry creature from Alpha Centauri, who might be green without having eyes capable of registering the fact. In contrast, there is a large proportion of humanity in which, as far as one can tell, mathematical ideas cannot be reproduced. Accordingly, the reproducibility of our mathematical ideas in the minds of small, furry creatures from Alpha Centauri is not certain (though most mathematicians would, if asked, regard it as very good). Mathematical ideas are precise in any mind that contains them.

The essence of formalism is precision. We have already seen that almost no mathematics is, as it is presented and understood, precise enough for formalism. A typical piece of mathematics contains both explicit precision (written in mathematical symbols), and latent or implicit precision (written in natural language) that can, by a fairly straightforward, very boring, but not mechanical process, be made explicit. A strictly formal proof has no implicit precision at all - everything is explicit. Unfortunately, non-trivial ideas (as opposed to lists of words) are dependent on implicit precision. *The most important objection to formalism is not that it is impractical, but that it does not allow the expression of ideas.* However, it is the explicit precision of formal mathematics that gives metamathematics its power. Ideas derived from concepts cannot be more explicitly precise than those concepts. As a body of mathematics develops, implicit precision must appear to produce powerful, general ideas. However, this implicit precision is only precision at all if one can see how, in principle, to make it explicit, so mathematics and metamathematics must start from a base of explicitly precise ideas. Metamathematics' demands are met by formalism. Mathematics' demands are met by axioms. It is for this reason that axioms are important (and see also Bourbaki[9]), not for formalism's sake.

I want to sum up and embellish these ideas in a diagram (Figure 1). The first row refers to the development of some body of mathematics. The arrow (which refers only to the first row) is a sort of arrow of time, defined by the inevitable increase of implicit precision. The section of the row adjacent to the arrow is well understood; the sections beyond it are not understood at all. It is credible that any explanations of these sections would rest on similar (neurological?) principles, in which case the linear row I have, could be regarded as a loop. I cannot help but speculate that, in this event, one might be able to do metamathematics with a vengeance. The final row refers to computer science, and will mean more to some readers than to others. The arrow is really only a pedagogical arrow; progress in this direction is relatively straightforward. Most mathematics, as indicated by the final phrase of the first row, proceeds against it. The further to the left of the row that novelty is introduced (the further a germinal idea may be carried upstream), the more significant it is (the more ideas it will produce downstream). The births of group theory, topology, and non-standard analysis require completely new sets of axioms. Most developments relate to more specialised but sophisticated ideas.

THE 4-COLOUR MAP THEOREM

The most topical question in the philosophy of mathematics is whether or not a certain proof of the 4-colour map theorem, given by Appel and Haken in 1976, is valid. The unusual feature of this proof is its reliance on calculations performed by computer, which are far too extensive ever to be checked by a human. The centre of this philosophical debate is the problem of how to clarify the notion of "proof" so as to definitely include or exclude this "proof by computer". This problem is placed in a new perspective by the proposed definition of mathematics. The virtue of a proof, or any mathematical discussion, rests in the precise ideas it contains. The computer proof contains so few new ideas as to make it almost entirely uninteresting, its only merit being that it does claim to establish the truth of the 4-colour map theorem. However, as I hope to have indicated, truth is a concept derived from, and therefore subordinate to, the precise idea. The latter implies the former, since flaws in proofs invariably arise through a lack of precision, but it is the precise idea that justifies mathematics.

-->>

	Axioms	Standard theorems	Building-block ideas of experts in a field	What will be mathematics in a month's time
	Explicit precision (word)		Implicit precision (idea)	
	Formalism		Platonism	
Hardware	Executable code	Assembler	High level language	

Figure 1

REFERENCES

1 Davis, P.J., and Hersh, R. The Mathematical Experience. Harmondsworth: Penguin, 1983

2 ------, "Platonism, Formalism, Constructivism" ref 1, part 7

3 Thom, R. "'Modern' Mathematics : An Educational and Philosophic Error?" American Scientist 59 pp 695-9, 1971

4 Davis, P.J., and Hersh, R. "The Formalist Philosophy of Mathematics" ref 1, part 7

5 Curry, H.B. Outlines of a Formalist Philosophy of Mathematics. Amsterdam: North-Holland, 1951

6 Davis, P.J., and Hersh R. "Pi and pi-hat" ref 1, part 8

7 Heyting, A. Intuitionism: An Introduction Amsterdam: North-Holland, 1956

8 Johnstone, P.T. Notes on Logic and Set Theory. Cambridge University Press, 1987

9 Bourbaki, N. "The architecture of mathematics" American Mathematical Monthly 57 pp 221-32, 1950
10 Nagel, E., and Newman, J.R. Godel's Proof. New York University Press, 1958
11 Cohen, P., and Hersh, R. "Non-Cantorian Set Theory" ref 1, part 5
12 Godel, K. "What is Cantor's Continuum Problem?" in Benacerraf, P., and Putnam, H. (eds) Philosophy of Mathematics : Selected Readings. (2nd edn) Cambridge University Press, 1983
13 Lakatos, I. Proofs and Refutations. Cambridge University Press, 1983
14 Kitcher, P. "On the Uses of Rigorous Proof" Science 196 pp 782-3, 1977
15 Davis, P.J., and Hersh, R. "Lakatos and the Philosophy of Dubitability" ref 1, part 7
16 Appel, K., and Haken, W. "The Solution of the Four-Color-Map problem" Scientific American 237 no.4 pp 108-21, 1977
17 Tymoczko, T. "The Four-Color Problem and its Philosophical Significance" Journal of Philosophy 76 pp 57-83, 1979
18 Teller, P. "Computer Proof" Journal of Philosophy 77 pp 797-803, 1980
19 Detlefsen, M., and Luker, M. "The Four-Color Theorem and Mathematical Proof" Journal of Philosophy 77 pp 803-820, 1980
20 Davis, P.J., and Hersh, R. "Why Should I Believe a Computer?" ref 1, part 8
21 ------, "True Facts about Imaginary Objects" ref 1, part 8

THE SOCIETY

Christopher J. Fewster (Chronicler)

The Archimedeans Year 1987-8 began in March 1987, on the kind of spring day that reminds you of winter, with our AGM. Thus began our lean time when plans were laid, plots were hatched, navels were contemplated and found in need of a good spring clean, and radical new philosophies discussed.

Within days, we were plunged into the depths of Oxford, on a snow-bound spring morning for a scavenger hunt with our Isis counterparts. The Invariants proposed we stay inside in the warmth - but we were made of sterner stuff. "We come from Cambridge", said the Vice-President and so we set forth in search of cellos, lawnmowers, McDonalds staff hats and American tourists with remarkable success (and, it should be said, legality).

As ever, many social events were held in the Easter term after the exams, including a croquet match against the Faculty and the presentation of a Lecturer of the Year award at our traditional Garden Party (which also featured a short recital given by our Barber Shop subgroup). There was also the curious affair of the mistaken punt-joust (for which the world is not yet prepared) and our punt trip, during which we encountered piracy and guerilla attacks on the High Cam and three more of our number became eligible for Dampers membership.

The more serious aspects of the Society have also flourished this year as our well-laid plans came to fruition. An *Alternative Guide* for first year mathematicians was published and distributed to freshers in October, the purpose of which was to disseminate useful information to help students through their first few weeks. A successful stall at the Societies Fair and an incredibly successful squash brought 190 new members to the Society, as a result of which our accounts are in a very healthy state, despite a reduction in the membership fee. This allows us to run some social events and lunchtime meetings at a slight loss, and give a better service to the membership. Our multifarious subgroups are thriving - there was even competition for the leadership of the Othello Subgroup, a matter eventually settled by a trial by ordeal. We have held several evening meetings, including one given by Dr. Singmaster which continued until 1am (!) and as I write, we still have three major speakers to come this term.

During the year, Dr. J.R. Partington of Fitzwilliam took over from Dr. P.T. Johnstone as our Senior Treasurer. We would like to thank Dr. Johnstone for all the work he has done on our behalf and also Dr. Partington for accepting this post, which is vital to the running of the Society.

The last major event of this year will be the Triennial Dinner on 20th February, which promises to be quite an occasion. This will be followed by the election of the new committee at the end of the term. We wish them the very best of luck.

So how can I sum up this year in a pithy conclusion? I am tempted to quote Macmillan and say "You've never had it so good!". In the end, though, I will quote that slithy tove , Mark Owen, and say "It was frabjous". So it was.

THE ECCENTRICITY OF MATHEMATICIANS

John F. Bowers (School of Mathematics, University of Leeds, Leeds LS2 9JT)

Ever since Archimedes, the great Eureka Streaker, leapt from his bath, mathematicians have been popularly regarded as eccentric. There is also a popular image of the physical characteristics of a mathematician, but this conflicts with the easily confirmed observation that "Mathematicians are strictly irregular", so it seems that a major eccentricity of mathematicians is that of not being as people expect them to be. Notwithstanding this, it is possible to draw at least a conventionalized portrait of a normal mathematician, and such a portrait appears in figure 1. The extent to which mathematicians differ from the norm is studied in Fudgett[6] and the exact nature of the investigation is indicated by figure 2, which plots the percentage of the mathematicians from a sample within each range of height.

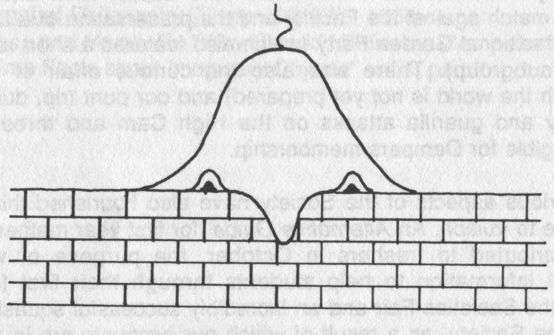


Figure 1. A normal mathematician.

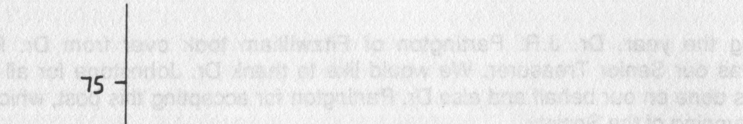


Figure 2. The heights of mathematicians.

That mathematicians are conventional but diverse in physical appearance should cause no surprise, because one possesses the standing of a mathematician on a purely mental feat, so, as a class, they are likely to share only intellectual, psychological, and emotional qualities. Consequently, the eccentricity of mathematicians must emanate from the mind because mathematicians contemplate the ultimate and unalterable truth of mathematics and thus have a true vision of the universe from an unusual angle. Indeed, to be among the few who appreciate such an esoteric truth is itself an eccentricity, but when combined with the intellectual power to generate further abstract truths, it is also the source of yet greater eccentricities.

This power of abstract thought is also useful in theology, which is probably why a surprising number of mathematicians have become bishops. The most recently appointed mathematical bishop is David Konstant (Leeds, 1985), and the earliest British example was Thomas Bradwardin (Canterbury, 1349), who had previously been Rector of the University of Oxford and whose mathematical work included the determination of the rules of proportion in arithmetic. Diogenes O'Rell summarized his career in the following words.

Thomas Bradwardine, in Britain born,
Oxford's great, God-witting guide
And blessed Bishop of Becket's bounds,
Of ratio's rules rightly wrote.

Gerbert (940-1003), who wrote a book on the use of the abacus, became an even more outstanding example of a mathematician and bishop in 999, when he was elected Pope and took the title Sylvester II.

To some extent, it is no longer possible to regard the set of mathematicians as a homogeneous population, because they now work in a number of largely independent subject areas. For example, it is no longer usual to work in both logic and cosmology. Indeed, the compartmentalization of mathematics introduces a danger that some subject areas (such as those which are amenable to linear techniques) might be studied in great depth but in complete isolation, while the remainder of mathematics is largely neglected. Diogenes O'Rell expressed his concern at this state of affairs as follows.

LINEARITY

In former times, all Mathematics lay
Like cowslips scattered on a gentle hill
Adjacent to a broad and winding way
Where wanderers could gather flowers at will.
And from this primrose path they did collect
By random methods all they chanced to find
But whether all their methods were correct
Or logicly arranged, they did not mind.

But nowadays such laxness is eschewed.
Instead we start from terms that we define
And then use means of matchless rectitude
Designed to treat equations of the line.
So down this straight and narrow path we send
Our Mathematics to its sterile end.

However, despite this specialization, because all mathematicians start by following similar courses of training and because there are interconnections between the subject areas which unite them all (though not necessarily any pair of them), it is still sensible to regard mathematics as a unit. Consequently, it is also sensible to use the term "mathematician" to mean anyone who works in some branch of mathematics, such as analysis or education, and hope to be able to prove general statements about all mathematicians, such as the following.

Theorem

All mathematicians are friends.

Proof. Any single mathematician is his own friend, because otherwise he would have no psychological peace and so could not work on mathematics, and thence could not be a mathematician.
Let us assume inductively that, for a positive integer n, every set of n mathematicians are mutual friends. Let us consider a set S containing n+1 mathematicians. In this set S, any set of n mathematicians are mutual friends, by the induction hypothesis, so the set S consists of mutual friends. In this way, we have deduced that any set of n+1 mathematicians consists of mutual friends. We may then deduce from the Principle of Induction that any finite set of mathematicians are mutual friends. As there is only a finite number of people in the world, there is only a finite number of mathematicians, so all mathematicians are friends.

In any society, tendencies towards eccentricity are controlled by the disapproval of the majority of the members. However, friends excuse even large divergences from normal behaviour and so form a society for the preservation of local eccentricities. Now, because of the esoteric nature of mathematics, mathematicians tend to associate mainly with other mathematicians and, because all mathematicians are friends, it follows that their eccentricities are not subjected to the usual controls. Further causes of such eccentricities have already been studied in Bowers[2,3] and pages 160-88 of Bowers[4]. One such cause is that the possession of mathematical ability implies the possession of other abilities because a mathematician needs to be able to engage in intellectual exploration, to recognize mathematical patterns, to generalize formulations, to create more abstract formulations and to utilize special cases of these abstractions. Consequently, all mathematicians share certain mental characteristics, as well as some educational background and a certain amount of work experience.

Undoubtedly, the greatest mathematical eccentricity is hard-mindedness, that is, the insistence that if a proposition is supported by a logically sound argument based on good evidence, then the proposition must be accepted. Perhaps we can redefine hard-mindedness as the vice of being right at the wrong time. An instructive example can be found in the story of the Dam Busters Squadron of the Royal Air Force. As usually told, the story has a villain in the form of a scientist who argued that the risks in attacking the concrete dam with the specially designed bomb were too high, that the destruction of the dam would not cause the intended strategic damage, although the destruction of the earth dam higher up the valley would do so, but the bomb would be ineffective against the latter dam. This last point was accepted because, as so often happens with inventions, the bomb had been invented first and the air raid had been planned afterwards in order to use it. However, the other objections were brushed aside and the air raid was carried out with great bravery. Although the concrete dam was destroyed, the official history of the air raid concluded that the scientist's objections were correct and the raid never had any chance of achieving its military objectives. Let us balance this account by considering the work of the *laisser-faire* economists in the

1840s. They insisted that the true principles of economics indicated that the market forces would produce a price level which, in the end, would be best for everybody concerned. They did not explain in what way their policies were best for the million or more people who starved to death in the British Isles, and especially in Ireland, as a consequence. In this way, even to this day, hard-minded politicians pursue their dire policies oblivious of the appalling consequences to others. Even if their hard-mindedness is incurable, mathematicians can reduce its effects by allowing for the existence of contrary evidence and by giving their opponents in argument plenty of time to accept the disputed proposition.

A second mathematical eccentricity is the overuse of logical analysis, brought about by much involvement with the linear form of a logical proof. This means that mathematicians are among the least able to determine how a conjuring trick is carried out. Suppose that the magician puts a billiard ball into his top hat and then, with typical insouciance, draws a rabbit out of it. An analysis of the event prompts questions about how the billiard ball and the rabbit were hidden in the hat, whereas the questions should really be about other hiding places and the methods of concealed transportation which were used. At least, this eccentricity lets a mathematician enjoy a well performed trick in an innocent way that is denied to the devious people who can work out how the trick is done.

A third eccentricity of mathematicians is deeply ingrained ergophobia. This starts harmlessly enough as a stylistic dislike of the excessive repetition of the word "therefore" (or, in Latin, "ergo") in all but the shortest proofs. To avoid this excess, the mathematician at first makes occasional use of variants such as "consequently", "thence", "thus", "hence", "this implies", etc., but soon goes on to eliminate the word "therefore" altogether. Such a fear of the word "therefore" represents fully developed ergophobia, which entails serious consequences. Suppose that a typically ergophobic mathematician encounters a problem which can be solved straightforwardly by about ten pages of arithmetic. Fear of such work will immediately set the mathematician to work to find a general method to solve such a problem without extended arithmetic and, in a few months, a theorem corresponding to this method might well be proved that allowed the problem to be solved in about half a page of working. In this way, ergophobia is the principal instrument of mathematical progress. Indeed, mathematics can be defined as the intellectual application of the principle of least action.

An eccentricity of which mathematicians are frequently accused is that of being absent-minded. Presumably, the charge is that mathematicians are not present-minded. In fact, a plea of guilty might well be entered to such a charge, because the universal nature of mathematical truth requires all mathematicians to be timelessly-minded.

A final eccentricity which is popularly supposed to inflict mathematicians is that of being prodigious. It is well known that even Isaac Newton was not a mathematical prodigy, but the belief continues because of the existence of undoubted mathematical prodigies, who are capable of mathematical work so well in advance of their ages that they strain the mathematical resources of their teachers. There are, however, three other kinds of supposedly mathematical prodigy. A second valuable kind is an *algorist*, who can manipulate mathematical formulae subconsciously and so produce suggestions for results without further explanations. The other kinds are arithmetical prodigies, some of whom are idiots, and those with visual intuition, which is a useful gift provided it is combined with some other suitable talent.

What are true mathematical prodigies like? Camilla Benbow[1] studied 100,000 intelligent children and noticed that the mathematically best 3000 contained twice as many boys as girls. Similarly, she found that twice the normal proportion of the precocious mathematicians were left-handed or suffered from allergy ailments. The proportion of girls in this sample contrasts sharply with the report[7] that only 10% of American Ph.D.s in mathematics are women. That report also recorded that two thirds of the women mathematicians in Chicago found that mathematics was a disadvantage to their social life, but that two thirds of the men made the same claim. According to the same survey, the women found that they needed to give more priority to their social lives than the men did. It was also reported that the women felt that they has received some discouragement during their careers, whereas the men felt they had not. This view is consistent with the investigation[8] by Rosie Walden and Valerie Walkerdine which found that the differences between schools was greater than the differences between boys and girls in tests for mathematical ability, but that some male-biased criteria seemed to be used in subjective assessments.

Another significant observation by Camilla Benbow[1] is that a large proportion of the mathematically precocious children suffered from myopia. Perhaps this is a contributory reason for mathematicians to be better at sedentary rather than active sport. The most notable sedentary sportsman among mathematicians was Emmanual Lasker, who was World Chess Champion from 1894 to 1921. However, mathematicians seem to have played chess for at least 850 years, as is clear from Omar Khayyam's[5] quatrain 49.

'Tis all a Chequer-board of Nights and Days,
Where Destiny with Men for Pieces plays:
Hither and thither moves, and mates, and slays,
And one by one back in the Closet lays.

Omar Khayyam is now remembered more for his poetry, such as the following quatrain (number 11), than for being probably the greatest mathematician of his day (about 1100), and for his considerable contributions to astronomy and to the solution of cubic equations.

Here with a Loaf of Bread beneath the Bough,
A Flask of Wine, a Book of Verse - and Thou
Beside me singing in the Wilderness -
And Wilderness is Paradise enow.

The list of mathematical poets includes Sir William Rowan Hamilton, Augustus De Morgan and Joseph Sylvester, whose long collaboration with Arthur Cayley on the study of algebraic invariants has been celebrated by Diogenes O'Rell as follows.

Arthur Cayley
Found new invariants daily.
He always sent the best t'
Joseph Sylvester.

Of course, some mathematicians enjoy poetry, just as many more enjoy music, because it can compensate for the lack of intrinsic emotion in mathematics.

A consequence of the poor sight of many mathematicians is that they are forced to view solid objects from many angles in order to appreciate their shape, and they also need to resort to topological methods in order to determine whether two solids are actually different. Diogenes O'Rell has characterized the most important factor in this geometrical aspect of topology as follows.

I'm in a ring but not in a ball,
I'm in a circle but not in a disc,
Even when there, I'm not seen at all,
So my presence is gravely at risk
Of being destroyed or completely ignored.
I can't be bought at the wholesale stores,
But I'm often so thoroughly bored
That you'll find me in the Henry Moores.
I'm the middle
Of a riddle.

However, it is not obvious that topological methods are adequate to make all the distinctions that are necessary in everyday life. For example, can a topologist tell the difference between a pillar box and an elephant's hind leg? For the purpose of this problem, the leg needs to be regarded as notionally detached from the elephant, and the pillar box should be regarded as a hollow cylinder which is closed at the top and the bottom and which has a slot on one side for posting letters.

In order to solve this problem, let us first simplify the form of the elephant's hind leg. To do this, we make the standard assumption among topologists that the materials of the leg (bone, flesh, skin, toenails, etc.) are entirely irrelevant, consequently we can roll the entire leg up into a tidy ball. At least, we can do so notionally. So is a pillar box topologically equivalent to a solid ball? If we push the top of the pillar box down and the bottom up, we transform the pillar box into two flat plates with part of a cylinder joining them. By straightening the sides and back and then pushing all the back continuously forward, we obtain a solid lump of cast iron which we can, notionally, roll into a ball. Therefore, we discover, a topologist cannot tell the difference between a pillar box and an elephant's hind leg. And, as any child can tell you, the moral of this story is that you must never ask a topologist to post a letter.

However, mathematicians can help each other to a far greater extent by swearing the following oath.

The Archimedean Oath.

The mathematician shall:

1. never teach anything that does not have a precise proof,
2. never publish any statement of which there is no written, detailed proof, and
3. never overfill the bath.

REFERENCES

1. Camilla Benbow, lecture to the American Association for the Advancement of Science, reported in New Scientist 110 (1986), part 1510, p. 24.
2. John F. Bowers, "Why are mathematicians eccentric?", New Scientist 100 (1983), pp. 900-3.
3. John F. Bowers, "Are mathematicians eccentric?", Journal of the Royal Society of the Arts 133 (1985), pp. 745-755.

4. John F. Bowers, Invitation to Mathematics, Basil Blackwell, Oxford, 1988.
5. Edward Fitzgerald, The Rubaiyat of Omar Khayyam, (first edition republished), Gay and Hancock, London, 1917.
6. I. Fudgett, "The subnormality of the normal mathematician's abnormality", Balonia Mathematica 19 (1975), 19-75.
7. New Scientist 95 (1982), p. 541.
8. Rosie Walden and Valerie Walkerdine, "Girls and Mathematics: from Primary to Secondary Schooling", Institute of Education, London, 1985; reviewed in New Scientist 108 (1985), part 1479, p. 17.

COLOURING PATTERNS OF SQUARES (OR, 101 USES FOR A VERY LARGE CHESSBOARD)

Graham Nelson (Selwyn College, Cambridge)

This article, while it does discuss a particular problem that was of interest to the author during his happy hours of commuting to London over the vacation, is also supposed to give some idea of some techniques used in graph theory, as applied to recreational maths. To quote one of Dr Hyland's example sheets, "this is not instructive but shows what can be done"...

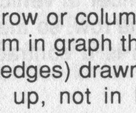
Take a (large) chessboard, and choose some arbitrary set of squares on it. If we want to colour in these squares, so that no two squares with the same colour are in the same row or column as each other, how many colours do we need? This is a typical problem in graph theory. A graph, curiously, is just a collection of points (vertices) with lines (edges) drawn between. We're usually only interested in the way the vertices are joined up, not in how they are arranged in three-space. We adopt the following standard notation:

K_n is the complete graph of order n, ie the graph of n vertices each of which is connected to every other.

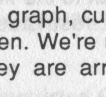
P_n is the path of order n, ie the graph which looks like a straight line with n vertices on.

C_n is the circuit of order n, ie the n-polygon with a vertex at each vertex.

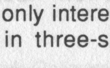
A vertex-subgraph of a graph G is a graph with some of the vertices of G, and all the edges of G which join them together.



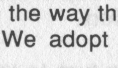
K_5



P_5



C_5



G_3

Now, define the graph G_n as follows: take an n-by-n chessboard and put a vertex on each square. Then join two vertices by an edge if and only if they are in the same row or column as each other. (This closely resembles the back of a home-built hi-fi when drawn out for $n>5$). A grid is a vertex-subgraph of G_n for some n. It's useful to measure the general complexity of a grid; we say that a grid is k-restricted if it has at most k squares in each row or column. 1-restricted grids are obviously rather trivial (they have just one vertex and no edges). We can classify 2-restricted grids as follows:

Theorem 1: Down to isomorphism, the connected 2-restricted grids are precisely the graphs P_n for $n \geq 1$, and C_n for n even.

Proof: Clearly the graphs mentioned can all be constructed as connected 2-restricted grids, so we just have to show that if G is such a grid, then it is one of the above graphs. Either G has a vertex connected to only one other, or it doesn't. If it does, then consider its neighbour. This vertex can have at most one other neighbour, because a neighbour has to be in its row or column, and we know there are at most two vertices in any row or column. By similar argument, we see that G is just a chain of points connected together, ie is P_n for some n. If it doesn't, then every vertex is connected to one other in the same row, and one other in the same column. Since G has only a finite number of vertices, we must have just a circuit graph, C_n for some n. Now pick any vertex, and observe that it connects via row to one other; this connects via column to a new vertex: which connects via row to another yet; and so on. If there are an odd number of vertices, we see that the final vertex in the circuit connects back to the first via the row, so that there are three vertices in the same row - but G is 2-restricted, which is a contradiction. Hence n is even.

We now turn to the colouring problem. Let k(G) denote the largest number of vertices in any row or column of G, so that G is k(G)-restricted but not (k(G)-1)-restricted. (Graph theorists sometimes call k(G) the "clique number"). It turns out that classical colouring techniques, dating back to the 1890's, will solve this problem completely:

Theorem 2 (the Grid Colouring Theorem): A k-restricted grid is k-colourable.

Proof: For k=1 this is easy - there's only one vertex, so colour it aquamarine and we are done. For k=2 we use Theorem 1; we can obviously easily colour the points of P_n or C_n with two colours as required. (Note that we do need n even in this last case!) Now suppose the theorem is false for some $k \geq 3$, and let G be the smallest grid which is k-restricted but not k-colourable. We get a contradiction by actually providing a colouring: pick any vertex v in G. Then take the grid of the vertex-subgraph of G formed by all the other vertices (ie, delete v). We can k-colour this since it is a smaller grid than the smallest which can't be k-coloured. So we do so. If the neighbours of v do not now use all k colours up, we can just colour v with one of the spare colours, so we are done. Suppose, then that the neighbours of v do use every colour. Since G is k-restricted, there are at most k-1 neighbours of v in the row, and k-1 in the column.

Claim: there is a pair of neighbours of v, one in the row and one in the column, which have (different) colours not used by any other neighbours of v. We call these neighbours x and y, and (wlog) they are brown and purple respectively.

Proof of claim: if not, then there are only as many colours used by the neighbours of v as there are neighbours in its column - but there are only k-1 at most, so all k colours have not been used - a contradiction.

We now apply the classical technique. Let K be the vertex-subgraph of G which has for vertices the brown and purple vertices of G. (K is usually called a "Kempe chain"). Now K is 2-coloured, so it must be 2-restricted; for if it had three vertices in any row or column, one could be brown and another purple, but a third can't have either colour and so can't be in K. We next show that x and y are in different components of K; suppose not, then there's a path between them in K; this path must have either odd or even length! It clearly can't have odd length since the colours purple and brown alternate along it, and so the two ends would have the same colour - a contradiction.

So it must then have even length. If this path is $p_1p_2p_3...p_n$, with $p_1 = x$ and $p_1 = y$, we have that p_1 and p_2 are in the same column as each other, p_2 and p_3 in the same row, ... p_{n-1} and p_n in the same column, but this means that p_{n-1}, p_n and v are all in the same column, so p_{n-1} is another purple-or-brown neighbour of v - a contradiction since by our construction, x and y are the only such neighbours. So we finally reach a contradiction to our claim that x and y were in the same component of K. Let K′ be the component of K which contains x. Now, we just change the colouring we already have so that all the brown vertices in K′ become purple, and vice versa; all the other vertices are unchanged. This is clearly still a legal colouring. But, now x and y are both coloured brown, since x has changed from purple to brown and y is unchanged - so v has no purple neighbour, and we can complete the proof by colouring v purple. This gives us a k-colouring, so that G wasn't a counter-example after all.

The (rather wordy) nature of this theorem disguises the actual simplicity of the solution, which is a pity. It is actually a construction for the colouring we want, but in more concise form:

Algorithm for Colouring a k-restricted Grid with k Colours:

1. Start by colouring all the squares grey, (not one of the k colours).
2. Find a grey square. If there are none left, stop.
3. Do the squares in its row and column use all the k colours? If not, give it one of the remaining ones and go to 2.
4. Pick a pair of colours such that one square in the column uses one - purple, say - one square in the row uses the other - brown, say - and no other square in the same row or column is either purple or brown.
5. Change the colour of the purple square to brown. Also mark the square with a dot.
6. If there are two squares in the same row or column as each other with the same colour, change the colour of the one without a dot in it from purple to brown or vice versa. Then mark this square with a dot, and go to 6.
7. If there aren't, colour the original square chosen purple, rub out all the dots and go to 2.

The observant reader will notice that this algorithm is simple precisely because the proof of Theorem 2 allowed arbitrary choice of the vertex v. The main virtue of this result is that it gives a "best possible" answer - after all, G is bound to need k(G) colours in the large row or column alone, so it's certainly economical to use exactly k(G). However, in his vainer moments, the author is pleased to observe that it contains the Four Colour Theorem (but only for a somewhat restricted class of grids!)

There are many other facts easily seen about grids, a few of which are quite interesting. The following is a set of challenges for graph theorists:

(i) Which graphs can be grids? (I have no general answer, though certainly complete graphs; the only trees allowed are paths; non-trivial bipartite graphs aren't allowed, and in general $K_{m,n}$ can only be the edge-subgraph of the grid K_{m+n}; the possible circuits have order 3, or any even order.)

(ii) When are grids isomorphic as graphs? (I conjecture that such may be transformed into each other by the elementary moves of row and column swaps. David Asher found an interesting invariant of these, but I have no solution as yet.)

(iii) Which grids are planar graphs? (Kuratowski's theorem isn't very helpful despite the results of (i) because of topological equivalents. However, at least we can observe that planar grids are 4-restricted; the converse is false. I conjecture that 3-restricted grids, at least, are planar.)

(iv) If you put a rook down on one of these sets of squares, when can it make a rook's tour, ie visit each square once and return to the start? (ie, when is a grid Hamiltonian? I conjecture that it is iff it is 2-connected - another "best-possible" result - and have a rather messy construction to try and show this by applying surgery to lots of small circuits, but don't yet believe it.)

(v) Are the only grids of edge-colouring class 2 (ie which require d(G)+1 colours to edge-colour) just the K_n for n odd?

I offer a box of Jaffa cakes (of any desired brand available in Cambridge) for a proof or counter-example of any of the above conjectures!

REFERENCES

The first three have sections on colouring problems - the first is especially recommended.

1 Saaty and Kainen, The Four Colour Problem: Assaults and Conquest, Dover
2 Bollobas, Graph Theory, Springer-Verlag (GTM series)
3 Yap, Some Topics in Graph Theory, Cambridge University Press (LMS series)
4 Hyland, Logic Example Sheet II, Cambridge 1987

WIT AND WISDOM 1

Ruth Lilley, Bob Dowling, Ian Redfern, and **Alex Selby**

As with all maths, when the real world starts to become a bit of a nuisance ... chuck it out. [Probability]

This does have physical applications. In fact it's all tied up with strings. [Oxford]

That's as precise as you need for statistics. Let me not knock it : it's not maths, but it's an interesting subject. [Probability]

I'm not sure if (a) I remember what I did last time, (b) I want to remember, or (c) I'm going to put it right.

This year's Honesty Prize goes to the natural sciences supervisor, who replied to a question with "Don't ask me. I'm not a mathmo."

I know you've got nought divided by nought, but that nought is smaller than this one ... I don't really expect you to believe that argument. [Linear Systems]

"Heuristically" means "handwaving".

Can anybody guess the context in which the following is correct? "This theorem is obviously proved as 13 equals 15."

No analyst ever draws pictures in public. [This Analysis lecturer followed this comment with a picture.]

There are Huppert tensors - they're pretty useless. [Dr. Huppert]

There's plenty of things in this course you could disagree with, but if you disagree with this one I advise you to leave the room now.

We can slip the d/dy past the integral sign and watch it curl up a bit.

We never mix up real and complex vector spaces - except perhaps we might.

The more I think about this, the more I realise this is a pretty stupid proof, but there you go. [His audience did.]

1/2 does not tend to nought as anything tends to anywhere.

This integral doesn't converge, but I'll evaluate it for you anyway. [Math. Methods]

The non-uniqueness is exponentially small.

J.S. Wilson is an infinite group theorist.

THE POWER OF MATHEMATICS

Lynn Arthur Steen (St. Olaf College, Northfield, Minnesota)

Mathematics is the enabling discipline of science and technology, the foundation of our information age. We live in a "minds-on" world created by the abstract theories of contemporary mathematics. Not just scientists but all citizens must now live and work in an environment increasingly influenced by advanced mathematics that is largely hidden from public view.

What little the public does know of mathematics comes from debate about the school curriculum. Mathematics instruction accounts for 10-15% of total instructional effort in education. Apart from English, mathematics is the most dominant single discipline in the entire curriculum.

It was not always so. Classical curricula devoted scant attention to any mathematics beyond Euclid. The explanation for why it came to be mathematics rather than, say, philosophy, that occupies this central position can be found in the scientific revolution: mathematics became *the* foundation discipline of science.

The roots of the scientific revolution lay in the introduction by Galileo of empirical methods to replace the speculative explanations that were part of the classical Greek tradition. Newton then created theoretical science by showing that empirical data can be explained by mathematical methods applied to basic axioms. Now, three hundred years after publication of *Principia*, we are entering a third scientific epoch - the age of computational science. This methodology, envisaged by the great Hungarian-American twentieth century mathematician John von Neumann and made possible by high speed computers, uses the results of theoretical science to simulate empirical reality: in this scheme, one calculates (rather than solves) the appropriate configurations that suggest how nature might act.

Computational Mathematics

Computational methods are changing the nature of science, the nature of mathematics, and the way each relates to the other. In the Newtonian tradition - which dominates science and mathematics curricula - the way one determines, for example, the time it takes for an object to fall from the top of a building is to solve the equation of motion to determine at what time the height becomes zero. In the von Neumann tradition, one would simulate the falling body by using the equation of motion as an instruction to the computer to calculate the height of the ball at each millisecond after it is dropped, and then observe (rather than solve) to determine when it hits the ground.

This is obviously a contrived example. But the idea is profound, especially when one applies it to project the flow of blood through an artificial heart valve or the trajectory of a hurricane as it approaches a coastline. In these more complex cases, one cannot hope to solve the equations of motion exactly, so computational solutions are the only appropriate strategy.

Computational mathematics makes possible tomography, a marvelous blend of science, mathematics, art, and black magic. Tomography is an abstract mathematical theory that enables computers to reconstruct visual images of the inside of an object - the human body, the mantle of the earth, or a nuclear reactor core - by calculations based on a succession of one-dimensional slice-like shadow images created by some type of radiation that has passed through the object. Since under some conditions different interior densities can produce identical sets of shadow images, this type of problem does not have a unique mathematically determined answer. Tomography is an example of what is called an ill-posed problem requiring a blend of science, mathematics and computing for effective solution.

Numerical methods have been part of applied mathematics for decades. What's different now is that these methods have been made effective by high speed computers, and thus have become part of routine science. Simulating reality inside a computer is often cheaper (and safer) than conducting thousands of experiments. The experiment that led to the Chernobyl accident is a case in point. Soviet engineers were attempting to discover whether at the moment of reactor shut-down the coasting turbines could supply enough power to carry the reactor from nuclear to diesel power. One of the American nuclear engineers who was part of the International Commission that investigated the Chernobyl accident said that such an experiment would never be conducted in the United States: we would use computer simulation, not real reactors for the study.

Whereas formerly only theoretical scientists needed to use sophisticated mathematics, now the routine practice of scientific investigation often requires techniques from computational science. As science and applied mathematics moves increasingly towards computational methods, so must the school and college curriculum if it is to reflect a modern perspective: it would be outrageous to perpetuate into the next millenium the out-dated idea that all important equations can be solved by simple formulas with integer coefficients.

Surprising Connections

As the relation of mathematics to science has changed, so have the internal dynamics of mathematics itself. Major discoveries from one area influence work in distant fields, often revealing surprising connections among widely separated parts of mathematics and science. It is important that students learn that the discovery of intrinsic intellectual connections is a crucial part of the conduct of modern mathematics. Two examples - operator algebras and fractals - illustrate the power of surprising connections to influence the direction and applications of mathematics.

Operator Algebras. The theory of operator algebras provides a good example of the subtle connections among diverse fields. It is an esoteric subject, one that is not usually taught until graduate school - although the notion of "operator" does appear naturally throughout the mathematics curriculum. An operator is an abstraction of an action: adding two numbers is an operation on the numbers; rotating a cube is an operation on the cube; integrating a function is an operation on a function. Specific operators are studied throughout the mathematics and science curricula. But operators as an abstraction are reserved for advanced mathematics courses.

The theory of operators arose in the early part of this century, motivated in part by the need to find mathematical models for quantum mechanics. It was in quantum mechanics that, for the first time, profound scientific examples emerged of non-commutative structures - of actions that did not satisfy the familiar commutative rule of arithmetic that $3 \times 5 = 5 \times 3$. The unexpected difference found when operators representing momentum and position were multiplied in different orders formed the basis for Heisenberg's uncertainty principle, and gave rise both to the mathematical subject of operator algebras and to a philosophical dispute, not yet resolved, about determinism and predictability in the physical world.

Operator theory flourished as a branch of functional analysis in the middle part of this century, reaching a level of maturity in which pure mathematical research was carried on with hardly any connection to its original roots in physics. Of course physicists continued to use and develop parts of the theory useful to them as well.

In recent years two major streams of thought converged to put operator theory once more at the forefront of mathematical research: a world-wide effort involving mathematicians from a dozen different countries led to the discovery of new types of operators, and to new methods for classification. This in turn enabled researchers to apply operator theory to geometry, which led just recently to the discovery of important relations between operator algebras and the classification of knots - a vexingly difficult problem that had previously defied all attempts at solution.

The latest chapter in this tale - so typical of the way mathematics is created and used - is that biologists are now using the new mathematical theory of knots to understand the mechanisms of DNA replication. DNA is an extremely long molecular fibre that normally is all scrunched up into a tight knot. Imagine a strand of DNA, all coiled up, that has replicated itself. Now try to imagine how the duplicate strands can pull apart. It's like the magician's trick of effortlessly separating two intertwined rope knots. From motivation in the uncertainty principle of quantum mechanics to esoteric research in abstract mathematics and then back to the knotting of DNA is an amazing albeit typical circle of ideas that illustrates the versatility, the fundamental character, and the essential role of mathematics in comprehending our world.

Fractals. The new theory of fractals provides yet another example of the power of surprise, revealing both a fascinating new geometry of intermediate dimensions and the expansive power of mathematics to provide models for science. Fractals illustrate better than almost anything else in recent years how modern mathematics provides intellectual bridges among diverse parts of science.

The public view of fractals are the amazing computer-generated color prints of mathematical forms that resemble artificial worlds. Hidden behind these beautiful pictures is a stunning confluence of several quite different strands of research that individually and collectively are transforming much of modern mathematics.

First there were investigations begun by Benoit Mandelbrot at IBM into what he termed fractional dimensions (whence fractals) - of objects like coastlines or rocky surfaces that wiggled so much as to place them mathematically at an intermediate dimension - more than 1 but not quite 2, or more than 2 and not quite 3. From various parts of science - for example, from population biologists who studied the seemingly erratic annual changes in harvests of ocean fish - came many similar investigations of the dynamics of simple processes repeated over and over again. (In the case of ocean fish, it was the annual cycle of hatching, harvesting, and death.) From engineering came theoreticians who were trying to understand photographics of turbulence, for example of air going over the wing of a high speed airplane. And from astronomy and physics came investigations of self-similar phenomena, for example, of the pattern of stars in the sky which looks statistically similar under many different degrees of magnification.

Gradually what emerged from these investigations was a common body of theory that showed how orderly processes, repeated sufficiently often, could lead to chaotic phenomena. Chaos, in this view, was not the antithesis of order - and therefore somehow beyond the reach of scientific theory - but a natural consequence of order. Chaos, it turns out, is order iterated beyond comprehension. Applications of these ideas quickly spread to all fields of science, and with the aid of high quality computer graphics, produced pictures of models of chaos and order never before seen.

Pictures of fractals help us to see the invisible and imagine the unimaginable. As a full grown organism represents an astonishing unfolding of the information encoded in DNA, so fractals display an immense variety of detail, all generated from relatively simple instructions. Recent research has shown that fractal geometry can be used to compress the data required to store and transmit photographs, since the appropriate fractal elements can regenerate details in the original picture without having to remember explicitly every tiny detail. Fractals thus serve as a metaphor for life, as a simulation laboratory for exploration of the encoding and unfolding of information. Perhaps someday they will be seen as more than just a metaphor.

Opportunities and Initiatives

Applications of mathematics to science, business, and engineering have extended the power of mathematical methods to new areas and new disciplines. Dozens of fields could be cited to document the vitality of contemporary mathematics. I want to highlight four particular areas as examples of the unity and applicability of mathematical research: computational statistics, mathematical biology, geometrical mathematics, and nonlinear dynamics. Each of these offers opportunities for initiatives in mathematical research and for stimulating interest among prospective students of mathematics. They epitomize the future of mathematics.

Computational Statistics. The statistical sciences study problems associated with uncertainty in the collection, analysis and interpretation of data. Not surpisingly, the increasing use of computers to record and transform data has generated a host of new challenges for the statistical sciences.

For example, analysis of data from electronic scanning devices (in tomography, in aircraft or satellite reconnaissance, in environmental monitoring) has produced an urgent need for statistical analysis for data with an inherent spatial structure. Research in this emerging field of spatial statistics employs a wide variety of mathematical, statistical, and computational techniques: problems of separating signals from noise borrow techniques from engineering; ill-posed scattering problems employ methods of numerical linear algebra; and smoothing of data requires statistical techniques of regularization. Underlying all this is the inherent geometry of the problem, which in many cases is dynamic and non-linear

Many applications of statistics (e.g., clinical data from innovative medical protocols) involve small data sets from which one would like to infer meaningful patterns. Now, however, computationally-intensive statistical techniques can use the limited available data to generate more data with the same statistical characteristics. By resampling the given data repeatedly, these so-called "bootstrap" methods generate millions of similar possible data sets which yield accurate approximations to various complex statistics. By comparing the value of statistics for the given sample with the distribution obtained for all possible samples, one can determine whether the observed values are significant.

Mathematical Biology. Nothing better illustrates the potential for mathematics in the biological sciences than the many traces of mathematics behind the Nobel prizes. For example, the 1979 Nobel Prize in medicine was awarded to Allan Cormack for his application of the Radon transform to the development of tomography and CAT scanners. The 1984 Nobel Prize in chemistry was awarded to biophysicist Herbert

Hauptman, President of the Medical Foundation of Buffalo, for fundamental work in Fourier analysis pertaining to X-ray crystallography.

Indeed, recent research in the mathematical sciences suggests dramatically increased potential for fundamental advances in the life sciences using methods that depend heavily on mathematical and computer models. Structural biologists have become genetic engineers, capturing the geometry of complex macromolecules in supercomputers and then simulating interaction with other molecules in their search for biologically active agents. Using these computational methods, biologists can portray on a computer screen the geometry of a cold virus - an intricate polyhedral shape of uncommon beauty and fascinating geometric features - and search its surface for molecular footholds on which to secure their biological assault.

Geneticists are beginning the monumental effort to map the entire human genome, an enterprise requiring expertise in statistics, combinatorics, artificial intelligence, and data management to organize billions of bits of information. Ecologists - the first mathematical biologists - continue to use the extensive theories of population dynamics to predict the behavior and interaction of species. Neurologists now use the theory of graphs to model networks of nerves in the body and the neural tangle in the brain. Cell biologists study the replication of DNA using the newly-discovered algebraic classification of knots. Epidemiologists monitor the spread of AIDS with techniques that blend innovative statistics with classical analysis. And, finally, physiologists employ contemporary algorithms applied to nineteenth century equations of fluid dynamics to determine such things as the effects of turbulence in the blood caused by cholesterol or swollen heart valves.

Geometrical Mathematics. Ever since Euclid, geometry has been one of the major pillars of core mathematics. After decades of decline (especially in mathematics teaching), the geometrical view in mathematics has undergone a renaissance, assisted both by the development of new theoretical tools and by the power of computer-based visual representation. In a very real sense, geometry is once again playing a central role on the stage of mathematics, much as it did in the Greek period.

Geometry claimed two of the three 1986 Fields Medals, which were awarded to Michael Freedman and Simon Donaldson for work in the geometry of four dimensional manifolds. By exploiting properties of the Yang-Mills field equations that reflect the wave-particle duality of matter, Donaldson showed that the differential geometry of four dimensional manifolds was vastly different than that suggested by their topological structure. Freedman provided the topological classification. Together their work yielded not only deep understanding of four dimensional manifolds, but the surprising insight that in four dimensions there are differentiable manifolds that are topologically but not differentiably equivalent to the standard Euclidean four dimensional space. Already, insights from this work have led to applications in string theory - the new super-symmetric theory of elementary particles - thereby providing fresh evidence of what Eugene Wigner called the "unreasonable effectiveness" of mathematics in the physical sciences.

Computer graphics provide a powerful new tool that extends geometrical techniques into many parts of mathematics. Computers - especially supercomputers - can calculate and display various mathematical structures in visual form, thereby enabling mathematicians to "see" the significance of abstract patterns that before could only be interpreted by formal means. For example, visual representations of solutions of differential equations often produce conjectures that open up whole new insights into the behavior of the system which the equations represent. Geometrical studies themselves regularly yield innovative challenges in the design of new algorithms and data structures, with spin-off benefits to applications in computer science (e.g., database systems and word processing) far removed from the original geometric problem. A newly launched Geometry Supercomputing Project involving geometers and computer scientists on both sides of the Atlantic linked by high speed computer networks is one example of the growing interaction of researchers in geometry with those in theoretical computer science.

Nonlinear Dynamics. Only in recent years have we been able to provide mathematical analysis of problems that are essentially nonlinear (e.g., turbulence in fluids). This analysis has been made possible by novel analytical methods, clever numerical simulation, and visual display on computer screens. Applications range from airfoil design to plasma physics, from oil recovery to studies of combustion.

Nonlinear dynamics has yielded many surprises, including long-term localized structures (e.g., the Red Spot on Jupiter), deterministic (rather than stochastically) generated chaotic motion (typical of some weather phenomena) and fractal patterns at the interface between fluids (e.g. displacement of oil by water). The mathematics of nonlinear dynamics involves a great deal of traditional analysis (especially differential equations), reinforced by iterative processes, automata theory, and fractal geometry.

Computer display of nonlinear phenomena makes visible patterns that would never have been noticed by analytic means alone. In research on dynamical systems, on the transition from order to chaos, and on the emergence of fractal shapes from smooth flows, computers are to mathematics what telescopes and microscopes are to science: they increase by a thousand-fold the portfolio of patterns that mathematicians can see and investigate.

The Newtonian revolution not only established mathematics as a paradigm for scientific reasoning, but it also established determinism as a paradigm for the behavior of physical systems. Nonlinear dynamics - a direct descendant of Newtonian mathematics - shows how ambiguity and uncertainty can arise in even simple deterministic systems, and how the onset of chaos itself can be predictable. In its power to change our Newtonian view of mathematics, nonlinear dynamics is as revolutionary as quantum mechanics; each breaks the bond of determinism and reveals entirely new structures that often defy what we have come to think of as common sense.

All One System

Despite appearances to the contrary, mathematical research is inextricably entwined with mathematics education at all levels, with science and engineering, and with political, economic, and sociological aspects of society at large. Educators and researchers, teachers and professors, mathematicians and scientists - we are all part of a single system of knowledge on which contemporary society depends.

The strength of the mathematical enterprise, therefore, depends not only on the internal merit of mathematical accomplishments, but on the vitality of the links that bind mathematics to education and to society at large. To insure this vitality will require that we

- Involve students in the practice of mathematics;
- Educate the attentive public;
- Explore fundamental issues in mathematics education;
- Ensure for all students equal opportunity for mathematical success;
- Invest in today's education to strengthen tomorrow's research.

What transforms these causes from empty rhetoric to concrete options are the opportunities for education and communication implicit in the advances of today's mathematical sciences - in such areas as computational statistics, mathematical biology, geometrical mathematics, and nonlinear dynamics. It is in the frontiers of mathematical science - not in current textbooks or today's classrooms - that one can find the innovative and intellectually rewarding options needed to transform education, to excite our youth, and to educate the public.

This is a cause that can unite researchers and educators in a common challenge: to let the power and beauty of mathematics speak for itself.

MATHEMATICS HAS A FRONT AND A BACK

Reuben Hersh (Department of Mathematics and Statistics, University of New Mexico)

In the famous book, *The Presentation of Self in Everyday Life*, by the U.S. sociologist Erving Goffman, there is a chapter called "Regions and Region Behavior". There Goffman introduces the concept of the "front" and the "back": regions to which the public is admitted, and from which it is excluded. In a restaurant, for example, the serving area is the "front"; the kitchen is the "back". In a theater, of course, the front of the stage is for the audience; backstage is for the actors, stagehands, props, and costumes. In front, the actors (waiters) wear costume (uniform); in back, they change clothes or rest in their casual dress. In general, the front is the region to which the public is admitted, where service is performed; the back is a region restricted to professionals, where preparations are made to provide service.

Goffman's contribution was to extend this concept of the "front region" and the "back region" from restaurants and theaters to all or almost all institutions of modern life. In the university, the classrooms and certain parts of the library are the "front" where the "public" (the students) are served. The Chairman's or the Dean's Office are the "back", where the products (classes and courses) are prepared "behind the scenes".

There is nothing sinister in this separation; it is a practical necessity. Goffman gives examples of the distress that can arise from blurring the line between "front" and "back". For instance, a gasoline (petrol) station whose customers feel free to wander into the parts department and help themselves to wrenches and hammers.

Goffman quotes Orwell (p. 121):

> "It is an instructive sight to see a waiter going into a hotel dining room. As he passes the door a sudden change comes over him. The set of his shoulders alters; all the dirt and hurry and irritation have dropped off in an instant. He glides over the carpet, with a solemn, priest-like air ... he entered the dining room and sailed across it, dish in hand, graceful as a swan."

The waiter who does this is performing automatically. If his split persona were brought to his attention, he would acknowledge its existence. But in the ordinary course of things, he just waits on tables. He is not conscious of putting on an act, or fooling anybody.

My purpose here is to point out that, like other social institutions, mathematics too has its "front" and "back", and to identify and describe them. It should be clear that now we are not speaking of "regions" in the literal, physical sense, as in dining room and kitchen. Mathematics is not necessarily associated with any particular physical setting; it is just a certain sort of activity. So its "front" and "back" will be particular kinds or aspects of mathematical activity, the public and private, or the part offered to "outsiders" (down front) versus the part normally restricted to "insiders" (backstage).

In this sense of the term, the "front" of mathematics is mathematics in "finished" form, as it is presented to the public in classrooms, textbooks, and journals. The "back" would be mathematics as it appears among working mathematicians, in informal settings, told to one another in an office behind closed doors.

Compared to "backstage" mathematics, "front" mathematics is formal, precise, ordered and abstract. It is separated clearly into definitions, theorems, and remarks. To every question there is an answer, or at least a conspicuous label: "open question". The goal is stated at the beginning of each chapter, and attained at the end. Compared to "front" mathematics, mathematics "in back" is fragmentary, informal, intuitive, tentative. We try this or that, we say "maybe" or "it looks like".

Observe that in all our examples the front is divided into subregions, of first, second, and even third class. A restaurant, for example, may include both a banquet hall and a snack bar. A theater has box seats, orchestra, and balcony. And the public for mathematics includes, among others, professional mathematicians themselves, graduate students, and undergraduates. The back is also divided, for efficiency and convenience, into subregions. In a restaurant, there are the domains of the salad chef, pastry chef, dishwasher, and so on. The reader can fill in the analogous divisions among working mathematicians.

The purpose of a separation between front and back is not just to keep the customers from interfering with the cooking. It is also to keep the customers from knowing too much about the cooking. Everybody down front knows that the heroine of the melodrama is wearing rouge. They probably don't quite know what she looks like without it. The diners know what's supposed to go into the ragout, but they don't know for sure what *does* go into it.

We can describe this state of affairs by saying that the front/back separation makes possible the preservation of a myth - whether it be the flavoring of the food or the beauty of the actress. By a myth we will mean simply taking the performance seen from up front at face value; failing to be aware that the performance seen "up front" is created or concocted "behind the scenes" in back. This myth, in many cases, adds to the customer's enjoyment of the performance; it may even be essential. More generally, a myth is a story that possesses a certain allegorical or metaphorical power. It is not literally true, but it survives while the generations pass by. Such, for instance, was the myth of the divine right of kings. Such are the myths of Christmas and Easter, and of course the corresponding myths of other religions.

Mathematics too has its myths. One of the unwritten criteria separating the professional from the amateur, the insider from the outsider, is that the outsiders are taken in (deceived), the insiders are not taken in. It would be straining patience to try to compile a complete dictionary of myths in mathematics. We list a few; enough to illustrate our point, and to enable the reader (as an exercise) to extend the list at pleasure. There is the myth of universality; the myth of unity; the myth of certainty; and the myth of objectivity. On a more specific or concrete level, there are the myths of Euclid, of Newton, of Russell, Brouwer, and Bourbaki. To present, describe and refute all of these myths would generate a thick volume. We content ourselves with some provocative comments; the reader can follow them up with the readings listed in the bibliography.

First, the myth of Euclid. This is discussed on pages 322-330 of *The Mathematical Experience*. The Euclid myth is defined there as the belief that the books of Euclid contain truths about the universe which are clear and indubitable. In view of the general availability of *The Mathematical Experience*, we need not here go into a detailed discussion of the Euclid myth. We merely point out that advanced students of geometry, and certainly professional mathematicians, are well aware that Euclid's axioms are unintelligible, his proofs incomplete, and his results limited to very restricted and special cases. Nevertheless, in secondary schools, in watered-down versions that fail even to mention his impressive achievements in solid geometry, Euclid continues to be upheld as the ideal model of pure mathematics and rigorous proof.

In a similar way, the plaster Newton created in the 18th century ("God said, Let Newton be - and all was light") is intact as a myth; the complex historical reality of Newton is almost unknown, even among the mathematically literate.

The myths of Russell, Brouwer and Bourbaki - of logicism, intuitionism, and formalism - have also been treated in *The Mathematical Experience*. Formalism is the subject (object?) of a beautiful diatribe in the preface to Lakatos' *Proofs and Refutations*. Therefore, in the hope of encouraging the circulation of *The Mathematical Experience*, we pass on to the more general myths on our list.

(1) Unity : There is only one mathematics, indivisible, now and forever. Mathematics is a single/inseparable whole.

(2) Objectivity : Mathematical truth or knowledge is the same for everyone. It does not depend on who in particular discovers it; in fact it is true whether or not anybody ever discovers it.

(3) Universality : Mathematics as we know it is the only mathematics that there can be. If the little green men (and women?) from Quasar X9 sent us their math textbooks, we would find again $A = \pi r^2$

(4) Certainty : Mathematics possesses a method, called "proof" or sometimes "rigorous proof", by which one attains absolute certainty of the conclusions, given the truth of the premises.

It would not be hard to find quotations to show that these beliefs are indeed widely held. Fortunately *Eureka* strives for entertainment, not pedantry, so we dispense with references. By calling these beliefs myths, I am not declaring them to be false. A myth need not be false to be a myth. The point is that it serves to support or validate some social institution; its truth is irrelevant, and most likely not determinable. Who can say, for example, that the doctrine of the divine right of kings is false? In the absence of a clear channel to the mind of God, this dogma can never be absolutely proved or disproved. But it was a useful belief, which in its time was credible, and served a purpose.

In a similar way, the unity, universality, objectivity, and certainty of mathematics are beliefs that support and justify the institution of mathematics. (For mathematics, which is an art and a science, is also an institution, with budgets, administrators, publications, conferences, rank, status, awards, grants, etc.)

Part of the job of preparing mathematics for public presentation - in print or in person - is to get rid of all the loose ends. If there is disagreement whether a theorem has really been proved, then that theorem will not be included in the text or the lecture course. The standard style of expounding mathematics purges it of the personal, the controversial, the tentative, producing a work that acknowledges little trace of humanity either in the creators or the consumers. This style is the mathematical version of "the front".

Without it, the myths would lose much of their aura. If mathematics were presented in the same style in which it is created, few would believe in its universality, unity, certainty, or objectivity.

Beliefs (1) through (4) are not self-evident or self-proving; they can be questioned, doubted, or rejected. Indeed, by some people they *are* rejected. Standard and "official" as these doctrines are, they are not taken so literally, so naively, by the backstage people. (A busboy or a stagehand is likely to be sceptical about the contents of the stew or the complexion of the ingenue.) Let us examine them critically, in order from (1) to (4), to justify our calling them myths.

From a backstage point of view, then, what about myth number one, unity? We see pure and applied mathematicians cooperating sometimes, but more often unaware of each other's work, and usually working to quite different standards and criteria. The pure may even declare that applied mathematics is not mathematics at all ("Where are the definitions? Where are the theorems?"). Or even worse, it is *bad* mathematics. (See Halmos' article. It is a landmark piece for having the courage to express an attitude, common but unspoken among "pure" mathematicians.) And even within pure mathematics, it is plainly visible at meetings of the American Mathematical Society that any contributed talk is understood by only a small fraction of those present at the meeting. The "unity" claimed in principle does not exist in practice.

As to myth number two, objectivity - yes, there is an amazingly high consensus in mathematics as to what is "correct" or "accepted". But alongside this, equally important, is the issue of what is "interesting" or "important" or "deep" or "elegant". These esthetic or artistic criteria vary widely, from person to person, specialty to specialty, decade to decade. They are perhaps no more objective than esthetic judgements in art or music.

And universality (myth number three) - who is to say? If there is "intelligent life" in Quasar X9, whatever we should mean by that, it might not be little green women and men. It might be blobs of plasma which we could not even recognize as intelligent beings. What would it mean to talk about their literature, or art, or mathematics? The very notion of comparing presupposes beings enough like us to make communication conceivable. But then the possibility of comparison is not universal; it's conditional on their being "enough like us".

And last of all, myth number four, certainty. Most of us are certain that 2+2=4, though we probably would find we don't all mean exactly the same thing by that equation. But it's quite another matter to claim equal certainty for the theorems of contemporary mathematics. Many of them have proofs which fill dozens of pages, which rely on other theorems whose proofs have not been rigorously checked by their users, which do not pretend to be complete but often contain such phrases as "it is easily seen" or "a standard argument then yields" or "a short calculation gives" and so on; moreover, more and more often, the paper will have several co-authors, not one of whom has carefully read the whole paper; and very possibly it will use the result of some calculations on a computing machine that none of the authors and possibly no living human being completely understands. Certainty, like unity, can be claimed only "in principle"; not in practice.

Myths, of course, need not be true. They need to be useful. Whatever the reason, it is clear that mathematicians *want* to believe in unity, objectivity, universality, and certainty, somewhat as Americans want to believe in the Constitution and free enterprise, or other nations in their Queen or their Revolution. But even while they believe, they know better. An important part of becoming a professional, in mathematics or anywhere else, is to move from the "front" to the "back". And part of this transition is to develop a less naive, more sophisticated attitude toward the myths of the profession. The leading lady needs her rouge. The stagehands know that she is the same actress they see behind the scenes with an ordinary, everyday face.

Reuben Hersh, Department of Mathematics and Statistics,
University of New Mexico,
Albuquerque,
New Mexico 87131

January 19, 1988

BIBLIOGRAPHY

Davis, P.J., "Fidelity in mathematical discourse: Is 1+1 Really 2?", American Mathematical Monthly, 78, 252-263 (1972)

Davis, P.J., and Hersh, R., The Mathematical Experience, Boston: Houghton Mifflin Company, 1981

Goffman, E., The Presentation of Self in Everyday Life, New York: Doubleday Anchor Books, 1959

Halmos, P., "Applied Mathematics is Bad Mathematics", in Mathematics Tomorrow, L. Steen, ed., Boston: Birkhauser, 1981

Lakatos, J., Proofs and Refutations, J. Worrall and E. Zahar, eds., Cambridge: Cambridge University Press, 1978

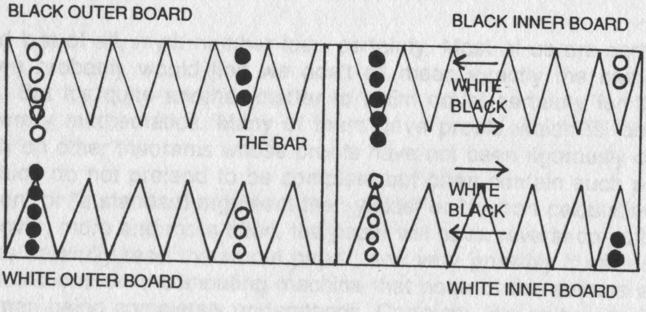
BACKGAMMON

Stuart Melbourne (Peterhouse College, Cambridge)

"Backgammon", I hear you say. A boring game played by weird people on a funny board. Well, don't worry about the board, it is just like it is so that we look clever playing on it. As for weird people, well, apart from our Treasurer, who has to be seen to be believed, we are as normal as anyone else in the University. As for the game being boring, the backgammon played by beginners (and NatScis) can be of the boring variety, but the game played properly can be spell-binding.

In the recent past (i.e. the last few decades - the game has been played in various forms for thousands of years) a new development has revolutionised the game. This is the doubling cube. It is essentially a gambling device. With it I can offer a "double" to my opponent to double the stakes of the game. If he accepts then we continue the game for double stakes and he then has control of the cube. He can thus redouble me if the game swings to his advantage - and the game can swing *very* easily, but I cannot redouble him. If he declines to take the double I win the game and we can go on to the next one. You would think that this would not greatly affect the game - as soon as one player has a small advantage he should double and his opponent should never accept. But it is easy to show (exercise for the keen!) that there is a range of probabilities for the outcome of a game where a double should not only be offered but also accepted. Neither player loses on his expected winnings. This opens up a whole new facet to the game, with rash doublers, slow doublers, etc. Control of the cube is also a great advantage - as long as your opponent cannot double you, you are in the game right to the close. Dramatic things can happen at the ends of these matches!

So much for the doubling cube. How exactly is the game played? Well, the rules are very simple.



THE INITIAL POSITION

Each player tries to move his pieces into his own inner board, using the values of two dice thrown each move. You may not land on any point occupied by two or more of your opponent's men. Any single opposition man may be hit (if the throw of one die moves one of your men onto the point held by this opposition man) by one of your men. If this happens, the "hit" man is removed from the board and placed "on the bar". Before your opponent can move any more of his men he must re-enter this man into your inner board. He is unable to re-enter his man on any point held by two or more of your men. Once *all* of your men are in your inner board you can begin to "bear off" - removing your men from the board in accordance with certain (simple) rules. The first player to bear off all of his men is the winner. If you win before your opponent has borne off *any* of his men, then you have won a gammon, worth double the points of a single game. If he still has men on the bar or in your inner board then you win a backgammon - worth triple the points of a normal game. The game sounds very simple, doesn't it? It is not.

Since you cannot land any of your men on a point held by your opponent with two or more men, a block of six such points (a prime) completely blocks in any men behind it. And a man "hit" and removed to the bar (where the whole society seems to end up on occasion) can prove very awkward - for either player. So, do you run for safety and try to get round the board as quickly as possible, or do you wait and hope to hit some of your opponent's men? If I knew the complete answer to that I would be a better player then I am. Perversely some players often aim to have their men hit so as to infiltrate men behind your lines and hit you back later, when they have most of their inner board filled and you cannot get your men back into play. Very sneaky, very dangerous, and often very effective. There is one more complication to contend with - a double on the dice counts as four, i.e. double three gives you four threes to play. This can have a wonderful effect or a disastrous effect, depending on your position at the time.

Perhaps I am still oversimplifying the game. Many positions are worthy of an analysis equivalent to that needed for chess, say. Those people who say that backgammon is a simple game of luck are usually rather poor players. Luck can have an effect in the short term, but over a period of only a few games the skill factor is usually decisive. The uncertainty of the dice just adds an extra spice to the game. A complete analysis of any game is impossible - the finest players in the world disagree on moves for some *opening* throws, never mind in the middle of a game. The beauty of backgammon is that you can analyse each position in whatever depth you require, from a trivial inspection to a complete probability picture for 2 or 3 moves ahead (an awesome task, with far more to think about than a 2 or 3 move plan in chess). With dice involved, over a short time you always have a chance to win - no matter how good your opponent. We can assure you of one thing - your belief in the laws of probability and statistics will be severely tested. Those of us who have played for any length of time will testify that Sod's Law is far more powerful. A pair of dice and a backgammon board can be a great leveller.

So, you've read the article - why not join the society? Come along and try us anyway. We welcome players of all standards, from absolute beginners to international tournament professionals. Our ladder system is designed to make the society competitive yet friendly. Who knows, you may get in the Varsity team to play at Oxford on February 28th this year. Quarter blues are awarded for outstanding play. We meet every Tuesday just after 8pm in the Rushmore Room at St. Catherine's College. Try us, you may find it habit forming!

For further information, contact :

Julian Hayward (President)	Peterhouse
Jason Charlesworth	Emmanuel
Stuart Melbourne	Peterhouse
Adam Jacobs	St. Catherine's

THE PROPOSTIONAL CALCULUS WITH DIRECTED GRAPHS

Martin Gardner and **Frank Harary**

Many formal logics can be represented by geometrical diagrams that are useful in two ways: they are visual aids that help students grasp the structure of statements in the logic, and in some cases the diagrams can be manipulated in such a way that theorems can be proved and problems solved as efficiently as with algebraic techniques. (For a history of logic diagrams see Gardner[1].)

The first good way to diagram the propositional calculus was by using Venn circles, proposed in 1880 by the British logician John Venn. He discussed his system at much greater length in Symbolic Logic[2], where he applied it almost exclusively to a class interpretation of the new Boolean algebra, especially to the diagramming of traditional syllogisms. Because Venn circles apply just as accurately to the binary relations of the propositional interpretation of Boolean algebra, it is curious that even today introductory textbooks of formal logic limit Venn diagrams exclusively to class-inclusion logic.

Figure 1 shows how a pair of Venn circles are shaded to represent each of the connectives of the propositional calculus. (Traditional symbols for the connectives are also shown.) Venn allowed the entire area outside his circles to represent the class consisting of the negations of all terms, but is best to confine this class to a small circle as indicated. This makes the region much easier to shade. Any diagrammed statement can be changed to its negation by exchanging the white and dark areas; the negated diagram is simply the photographic "negative". Three circles will handle all binary relations that concern three or fewer terms, each circle standing for a proposition that must be either true or false. The system is readily extended to diagrams for four or more terms by using closed curves other than circles, and to matrix forms proposed by John Marquand, Lewis Carroll, and others.

A method of diagramming the propositional calculus with undirected networks was proposed by Gardner[1] in the late fifties. In the early sixties Dr. Garrit M. Mes, a Dutch-born surgeon at the Medical Center, Krugersdorp, Africa, improved the network method by adding arrows to its lines. His system was never published, though it is briefly mentioned in Gardner[1]. In 1977 Frank Harary independently thought of the same technique. The purpose of this paper is to explain the method. The authors believe it is a useful educational device. Not only does it diagram statements in the propositional calculus in a way that permits the efficient solving of elementary problems, but it solves them in a way that Charles Peirce liked to call "iconic" - that is, in a manner that strongly resembles the formal structure being analysed. Moreover, it makes use of diagrams closely related to the networks of logic circuits in today's computer chips.

The digraph system has another great advantage over Venn diagrams. After diagramming the premises of a problem in the propositional calculus, using Venn circles, it is impossible to distinguish the shading of one premise from the shading of another. This makes it extremely difficult to experiment with the problem structure by altering its premises to see what the change entails. In the digraph system, as we shall see, each premise has a diagram isolated from the others. This makes for great ease in exploring the total structure of the problem; seeing how it changes when any premise is removed or altered, or new premises added.

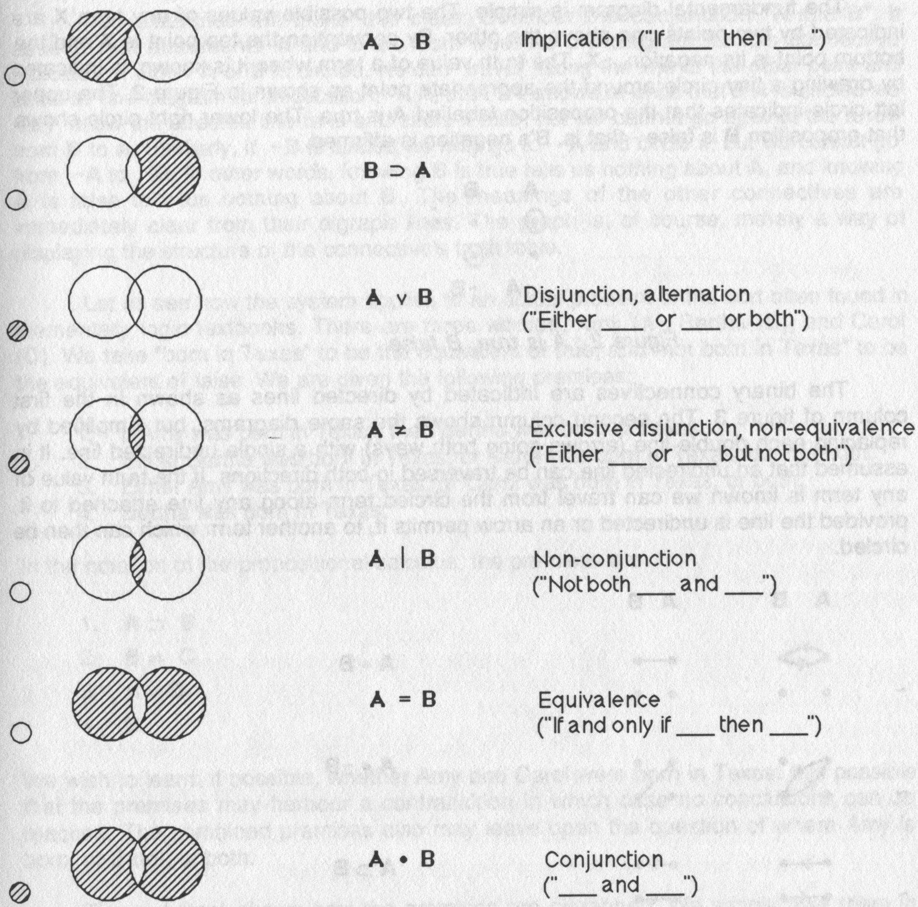

A ⊃ B
Implication ("If ___ then ___")
B ⊃ A
A ∨ B
Disjunction, alternation
("Either ___ or ___ or both")
A ≠ B
Exclusive disjunction, non-equivalence
("Either ___ or ___ but not both")
A | B
Non-conjunction
("Not both ___ and ___")
A = B
Equivalence
("If and only if ___ then ___")
A • B
Conjunction
("___ and ___")

Figure 1 : Venn diagrams for binary connectives in the propositional calculus.

The fundamental diagram is simple. The two possible values of any term **X** are indicated by two points one above the other. By convention the top point is **X** and the bottom point is its negation, ~**X**. The truth value of a term when it is known, is indicated by drawing a tiny circle around the appropriate point as shown in Figure 2. The upper left circle indicates that the proposition labelled **A** is true. The lower right circle shows that proposition **B** is false - that is, **B**'s negation is affirmed.

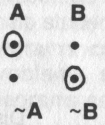


Figure 2 : A is true, B false.

The binary connectives are indicated by directed lines as shown in the first column of figure 3. The second column shows the same diagrams, but simplified by replacing each double line (arrows going both ways) with a single undirected line. It is assumed that an undirected line can be traversed in both directions. If the truth value of any term is known we can travel from the circled term along any line attached to it, provided the line is undirected or an arrow permits it, to another term which can then be circled.

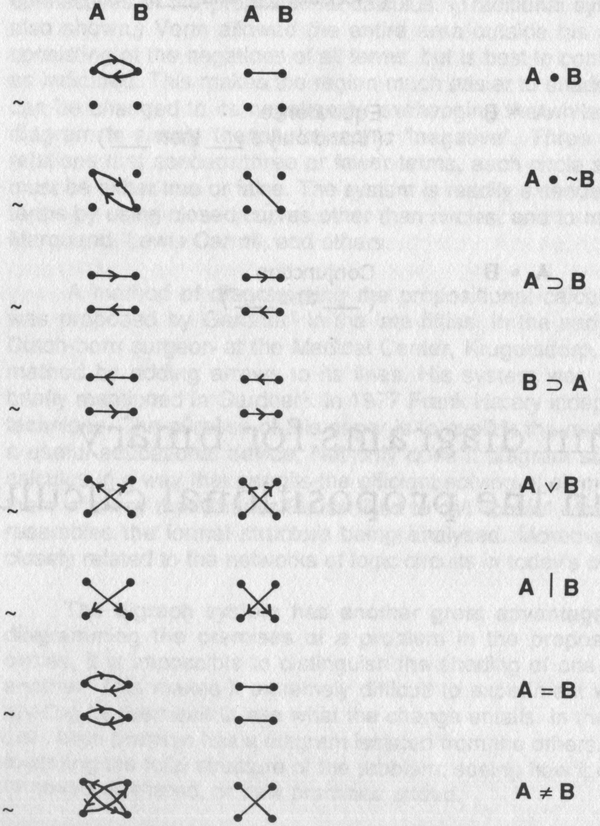


Figure 3

Two examples will make this clear. Consider the conjunction "**A** and **B**". It asserts that statements **A** and **B** are both true. As the diagram for this connective indicates, if either **A** or **B** is circled, we can "travel" along the line to the other point and circle it. The diagram for implication, "If **A**, then **B**", shows at once that if **A** is circled we may follow the directed line to **B** and circle it also, but we cannot go against the arrow from **B** to **A**. Simliarly, if ~**B** is circled we may go to ~**A** and circle it, but we cannot go from ~**A** to ~**B**. In other words, knowing **B** is true tells us nothing about **A**, and knowing **A** is false tells us nothing about **B**. The meanings of the other connectives are immediately clear from their digraph lines. The graph is, of course, merely a way of displaying the structure of the connective's truth table.

Let us see how the system applies to an actual problem of the sort often found in elementary logic textbooks. There are three women, Amy (**A**), Bertha (**B**), and Carol (**C**). We take "born in Texas" to be the equivalent of true, and "not born in Texas" to be the equivalent of false. We are given the following premises:

1. If Amy was born in Texas, then Bertha was born in Texas.
2. Either Bertha was born in Texas, or Carol was born in Texas, but not both.
3. Either Amy was born in Texas, or Carol was born in Texas, or both.
4. Bertha was born in Texas.

In the notation of the propositional calculus, the premises are:

1. $\mathbf{A} \supset \mathbf{B}$
2. $\mathbf{B} \neq \mathbf{C}$
3. $\mathbf{A} \vee \mathbf{C}$
4. $\mathbf{B}$

We wish to learn, if possible, whether Amy and Carol were born in Texas. It is possible that the premises may harbour a contradiction in which case no conclusions can be reached. The combined premises also may leave open the question of where Amy is born, or Carol, or both.

Figure 4 (top) shows how the premises are digraphed. We simply take them in order. After the three graphs are drawn for the first three premises, and **B** is circled (on the basis of the fourth premise), all identical terms are joined by undirected lines to make a single connected digraph.

We start exploring it at **B** because we know **B** is true. We cannot travel west along the line from the first circled **B** because the arrow prohibits it, but we can go from the second circled **B** to a ~**C**. We circle ~**C** as shown in Figure 4 (bottom). From ~**C** we can travel to another ~**C** and circle it. From there, we can go along the directed line to **A** and circle it. From **A** we go to the leftmost **A** and circle it. The path leads back to **B**. The digraph now looks as shown. We see that all the **A**s, **B**s, and ~**C**s are affirmed. We conclude that Amy and Bertha were born in Texas, Carol was not. There is no way to continue exploring the graph. No contradiction (affirming both a term and its negation) was encountered, therefore the problem is solved. If premise 4 had asserted that Bertha was not born in Texas (~**B**), a similar exploration of the digraph would have affirmed ~**A**, ~**B** and **C**.

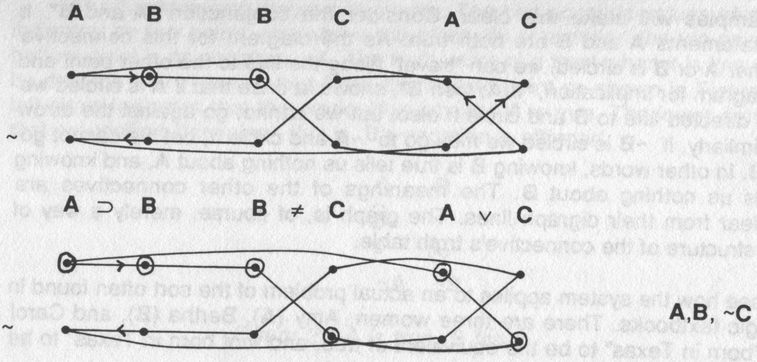


Figure 4 : Solving Problem 1 with digraphs.

It is easy to see how the technique can be extended in a chain to take care of any number of binary premises, and involving any number of terms. There are techniques for using such digraphs to handle compound statements (statements with parentheses) but, as is the case also with Venn circles, the system becomes too complicated to be of much interest.

If the premises do not affirm the value of any single term, they may still permit determining the value of one or more terms. In such cases one must explore the digraph by making assumptions about values to find out if any assumptions can be eliminated by a contradiction. To see how this works consider the following problem.

A woman may or may not be Ann (**A** or ~**A**), may or may not be beautiful (**B** or ~**B**), and may or may not be clever (**C** or ~**C**). We are told:

1. The woman is either Ann or she is clever but not both.
2. She is not both beautiful and clever.
3. She is either Ann or beautiful, or both.

In notation:

1. $\mathbf{A} \neq \mathbf{C}$
2. $\mathbf{B} \mid \mathbf{C}$
3. $\mathbf{A} \vee \mathbf{B}$

What can be deduced about the woman?

The digraph is shown in Figure 5. To explore it, we start with any assumption we please. Let's assume the woman is Ann. We circle the two **A**s. Traversing the lines allows us to circle two ~**C**s. She is not clever. No contradiction is encountered, but we have learned nothing about whether Ann is beautiful. We try again, erasing the circles and circling ~**A** to see what follows if we assume the woman is not Ann. Following the lines quickly leads to contradictions - forcing us to affirm all the terms and their negations. We need go no further. The premises are consistent only with the assumption that the woman is Ann, who is not clever. Whether she is beautiful is undecidable.

If we assume that the woman is beautiful, we find that she must be Ann, and not clever. But we can also assume she is not beautiful, and hence find that she is Ann and not clever. If we assume she is clever, the digraph leads into contradictions. If she is not clever, the digraph shows she must be Ann, with her beauty undecided.

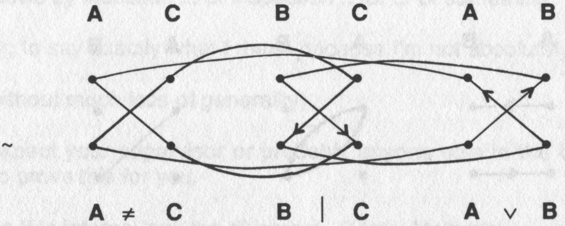


Figure 5 : Solving Problem 2 with digraphs.

It goes without saying that, as with all good logic diagrams, the digraph provides a simple way to prove tautology. If two statements are identical, their digraphs will be identical. For example, De Morgan's well-known pair of laws asserts that ~(**A•B**) (which is the same as **A | B**) is equivalent to ~**A** ∨ ~**B**; and ~(**A** ∨ **B**) is equivalent to ~**A** • ~**B**. The equivalences are obvious when we digraph each side of either law and see that the digraphs are identical.

To change the value of a single term **X** or ~**X** in a binary relation, imagine that the points at positions **X** and ~**X** are exchanged, carrying with them any lines that are attached. Think of a line as an elastic string, one end of which moves with a moving point while its other end remains fixed. Figure 6 shows, for example, how **A** ⊃ **B** is changed to **A** ⊃ ~**B**.

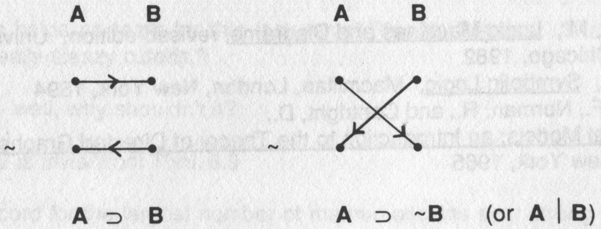


Figure 6 : Negating a term.

To change the digraph of an entire binary relation to its negation, first change each undirected line to a double line with arrows going both ways. Take the front end of each line and move it to the term's other point - that is, move it either up or down. Finally, if the result is a double line, replace it with an undirected one. Figure 7 shows an example. We want to negate **A** ⊃ **B**. When the front ends of the two directed lines are switched, we get the middle digraph. The double line with arrows going both ways can now be replaced with an undirected line as shown on the right. We see that the negation of **A** ⊃ **B** is **A** • ~**B**. In words, if it is false to assert that **A** implies **B**, then **A** must be true and **B** must be false.

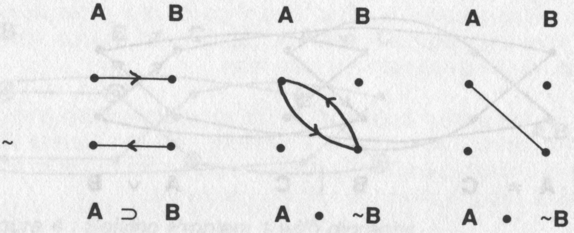


Figure 7 : Negating a connective.

The logic digraphs can be modified and extended in many ways that suggest further study. What is the best way to digraph compound statements? Can digraphs handle class-inclusion logic with existential quantifiers such as the syllogisms premise "some **A** is **B**"? Can digraphs be applied to other logics such as the logic of strict implication and other modal logics? Can they handle multivalued logics by adding more points to each term and perhaps using different colours for the lines to indicate the possible values of the connectives? Can digraphs be applied to fuzzy logic?

The authors would be pleased to hear from anyone who investigates these questions, or who introduces digraphs in elementary logic courses and cares to report on how the students responded.

REFERENCES

1. Gardner, M., Logic Machines and Diagrams, revised edition, University of Chicago Press, Chicago, 1982
2. Venn, J., Symbolic Logic, Macmillan, London, New York, 1894
3. Harary, F., Norman, R., and Cantright, D., Structural Models: an Introduction to the Theory of Directed Graphs. Wiley, New York, 1965

WIT AND WISDOM 2

Ruth Lilley, Bob Dowling, Ian Redfern, and **Alex Selby**

And that follows by induction ... or inspection ... or ... or something. [Logic]

I'm not going to say exactly what I mean because I'm not absolutely certain myself.

WMLOG (without much loss of generality).

You can't expect your supervisor or probably anyone else in the University except me to be able to prove this for you.

... as long as this integral isn't too divergent. [Math. Methods]

For x=0 the integral doesn't converge but I can fudge it for you if you like. [Math. Methods]

This is a relativistic situation because light moves at the speed of light. [Quantum Mechanics]

A lecturer recently started to use RUNES in his course! His justification: "I need an immediately distinguishable character... so I'll use something that no-one will recognise."

My mnemonic to help remember this is "Charges measure $\underline{E}$, currents measure $\underline{B}$".

I might be wrong - I'm not, of course, but in principle I might be.

Hold on a second - I need another axiom. [Halfway through a proof - the axiom was 0 does not equal 1].

Any theorem in Analysis can be fitted onto an arbitrarily small piece of paper if you are sufficiently obscure.

Renormalisation holds no fears for this lecturer of Plasma Physics: "...and divergent integrals need really sleazy cutoffs."

Why? Because - well, why shouldn't it?

Proof of Thm. 6.2 is trivial from Thm. 6.9

The previous record for the largest number of mathematicians ever strung together in a single string was six, occurring in the Liouville - Green - Wentzel - Kramer - Brillouin - Jeffreys Method of asymptotic expansion. This is no longer the case. In his book 'Mathematical Methods of Classical Physics', V.I.Arnol'd calls a result the 'Newton - Leibniz - Gauss - Green - Ostrogradskii - Stokes - Poincare Formula'. What he actually means is the combined Stokes' Formula and Gauss Divergence Theorem. This gives us a present record of seven. Can you beat this?

A GENERAL SURVEY OF THE MATHEMATICAL PHILOSOPHY OF IMMANUEL KANT

Robert K. Clifton (Department of the History and Philosophy of Science, Cambridge)

1. INTRODUCTION

On examining literature on the philosophy of mathematics one might conclude that the discipline is a fairly recent one beginning predominantly at the end of the nineteenth century with the writings of such thinkers as Frege, Peano, and Russell. Furthermore, it would not be uncommon to conclude that the subject appears quite inaccessible without the aid of a more specialized training in symbolic logic and higher mathematics. In the course of researching this article I have come to find that the above are misconceptions; that not only are the intellectual roots in this topic found as far back as classical Greek civilisation, and considerably more in the foundational groundwork of the great thinkers of the "Age of Reason", but that the fundamental issues in the field are more appropriately analysed from a general rather than a highly technical and esoteric point of view. With this in mind, I have chosen to survey the particular view of mathematics put forth by Immanuel Kant, philosopher and thinker of the Enlightenment period. Although philosophy of mathematics has advanced far beyond the point of Kant's Critique of Pure Reason, the main source of his mathematical conceptions, I feel that the Kantian influence on modern foundations research is profound enough to discount Russell's statement calling Kant 'a misfortune'. Rather, I take the more humble and realistic approach best enunciated by Paton (1936, I, 56) in his book Kant's Metaphysic of Experience:

> I believe myself that if we can penetrate, even imperfectly, into the argument of the Kritik, we shall find something other than a pedantic old professor armed with an external architectonic, incompetently tacking together old notes of what he used to think. We shall find instead a powerful and penetrating intellect struggling and twisting towards its goal.

A proper survey of Kantian ideas of mathematics and reality cannot be done without a philosophical and historical context in which to formulate these ideas. As well, I will attempt to highlight some standard objections to these ideas and, where possible, insert my own commentary. And to further underscore the importance of the substance of Kantian ideas an examination of modern Kantian influences will be given. My desire is that the main concepts of Kantian mathematical philosophy, although highly controversial and oftentimes confusing, will be seen to be influential for modern philosophy of mathematics.

2. KANT'S PHILOSOPHICAL AND HISTORICAL CONTEXT

Born in 1724, Kant was highly influenced by Leibniz and the theories of Newton and early in his life developed a high regard for mathematics. As he became more interested in Hume's empiricism[1] he strove to compromise it with Leibniz whom he knew and admired. As Maziarz (1950, 98) puts it:

---

[1] 'Empiricism' is the school of philosophy stressing sense experience of external physical reality ('a posteriori' knowledge) as the ultimate source of all knowledge of the world. This contrasts with 'rationalism' which emphasizes knowledge of our world as stemming from reason as derived from the mind alone ('a priori' knowledge).

> Kant aimed, then, to name the constitutive elements of mathematics and science whose certainty was unquestionable and to define the value and sphere of metaphysics[2] by properly articulating the extreme of rationalism as typified in the Leibniz-Wolff philosophy with that of Hume's empiricism.

This led to his general philosophy of transcendental criticism in which the respective roles of reason and experience were criticized and reunited in a 'transcendent synthesis'.

The second period of Kant's thinking, during his professorship of Logic and Metaphysics at the University of Konigsberg (1755-1770), took Kant deeper into Humean empiricism with his first major essay entitled "Is Metaphysics Capable of a Certainty Equal to Mathematics?" In this paper mathematical certainty was portrayed as a process of synthetic construction of axioms and definitions into concepts via a special 'intuition'. In 1770 Kant wrote the paper "A Dissertation on the Form and Principles of the Sensible and Intelligible World" in which he maintained "sensitive cogitations are representative of things as they *appear*, intellectual, however, of things as they *are*" (Lindsay 1934, Ch.2, *italics mine*). At this point, the distinction between sensible and intelligible knowledge was drawn; sensible knowledge exploring the phenomenal world by way of sensations induced through intuition and intelligible knowledge exploring the 'real', noumenal world 'behind' the phenomena. These 'worlds' roughly corresponded to the physical and metaphysical levels of reality respectively.

Briefly, Kant's first of three major works, entitled Kritik der reinen Vernunft (Critique of Pure Reason) and written in 1781, enquired about the nature of judgements and their origin and role in acquiring certainty. Although for the most part this was a revolutionary work, Kant continued to show his unquestioning acceptance of Aristotelian deductive logic and Euclidean geometry in the spirit of his scientific predecessors. In Kant's second critique, Kritik der praktischen Vernunft (Critique of Practical Reason), he attempted to show how the noumenal truths that he stripped from the province of speculative reason (freedom of the will, immortality of the soul, God's existence) arose. In his final critique Kritik der Urtheilskraft (Critique of Judgement), he attempted to synthesize his earlier critiques by sketching the relations between pure reason and practical reason. In all, Kant's quandary in the light of his time is best summed up, again, by Maziarz (1950, 103):

> Kant had begun with this problem: the miserable plight of metaphysics as alternately subjected to the strains of dogmatism or empiricism and the contrasting success of mathematics and science. History, he saw, revealed that applying mathematical and scientific procedure to the spiritual world yielded barren, sceptical results. The only conclusion he could validate was to credit mathematics and science with knowledge of the appearances of sense, and to deny the possibility of speculative, metaphysical knowledge of the nature of noumenal things.

It is evident, also, that Kant must have had a profound influence upon the chronologically close following positivistic movement whose sole goal was to make all forms of knowledge mathematical and scientific. Indeed Auguste Comte (1798-1857) canonized this trend following Kant's lead.

---

[2] The study of metaphysics is roughly the study of the reality lying "behind" physical reality.

3. OVERVIEW OF THE CRITIQUE OF PURE REASON

The bulk of Kant's mathematical philosophy can be found in his Critique of Pure Reason and for this reason detail will only be explicated concerning it. It is difficult to pick out Kant's philosophical assertions on mathematics and understand them properly without first looking at his general philosophy of reality which is intimately connected with his mathematical views. At this time, then, a first pass through the general ideas and structure[3] of the Critique[4] will be made in order to set the context for the specific statements Kant makes concerning mathematics in his work.

As mentioned previously, this first critique was concerned with the nature, origin, and role of judgements and the means by which they attain truth. Note that knowledge, for Kant, consisted in declarative judgements expressed in simple subject(A)-predicate(B) form, where both A and B denoted or signified a concept in order for the whole to be meaningful. Kant's general thought was that the current distinction, to his time, between necessary and contingent truths[5] was inadequate and although he felt Leibniz and Hume were closest to the truth they had failed to distinguish analytic and synthetic from a priori and a posteriori judgements. To Kant the relation between the subject(A) and predicate(B) of a judgement was either analytic or synthetic. Kant (1933, 55):

> Either the predicate B belongs to the subject A, as something which is (covertly) contained in this concept A, or B lies outside the concept A, although it does indeed stand in connection to it. In the one case, I entitle the judgement analytic, in the other synthetic.

What Kant meant by 'a priori' knowledge was not the usual (Leibnizian) meaning, that is, 'innate' knowledge gained prior to experience. 'A priori' knowledge, for Kant, begins with experience but does not directly arise from experience. (This was one of Kant's attempts to mediate empiricism with rationalism (see note 1)). By 'a posteriori' knowledge, Kant held to the traditional meaning, that is, knowledge gained from experience itself.

For Kant there were four logically possible judgement types. 'Analytic a priori' statements and 'synthetic a posteriori' statements corresponded to what Hume termed knowledge of idea relationships (trivial in Kant's opinion) and knowledge of matters of fact respectively. The possibility of 'analytic a posteriori' statements was denied. All other knowledge, mathematics being the prime example, fell into 'synthetic a priori' knowledge. Although he did not prove this he attempted to show how this type of knowledge was possible through his theory of perception.

In this theory of perception, contrary to his contemporaries, Kant held that the mind was not a mere passive recipient of stimulus but literally added something to the perception and reasoning processes. The Critique consisted of three parts: The Transcendental Aesthetic (nature of the double contribution of the mind to sense perception), The Transcendental Analytic (the contribution to pure understanding), and

---

[3] For further commentary see Baum (1973, 212-215) and Maziarz (1950, 100-102).

[4] Henceforth for brevity Kant's first critique is referred to simply as the Critique.

[5] Necessary truths are judgements that must be true under all conceivable circumstances, such as "Mathematicians study mathematics". Contingent truths are judgements like "All mathematicians are Archimedeans" which *need not* be true and depend upon other circumstances.

The Transcendental Dialectic (the contribution to pure reasoning). Kant, as previously noted, separated perception into the phenomena (stimuli) perceived and the noumena ('things-in-themselves') which create the sense stimuli. This was not scepticism but, on the contrary, cleared the way for faith in God (as he attempted to show later in the Critique of Practical Reason). He claimed the processing of raw sense data (the 'manifold of intuition') was accomplished by the understanding but before this the mind added the two forms of intuition - space and time. Contrary to Newtonian belief, space and time could not be viewed as independently existing entities or conceptual relations. This account, according to Kant, could be the only reasonable explanation of the possibility of a priori mathematical knowledge. The mind necessarily applied space and time to experience, forcing mathematical truths to be not only universal but necessary a priori, raising them to the level of mathematical judgements.

How was it possible for mathematical knowledge to be *synthetic*, according to Kant? It is here that his ideas become similar to those of John Locke. Locke believed mathematical knowledge to be universally necessary and non-trivial or synthetic a priori. So also did Kant, however for Locke the 'a priori-ness' of mathematical statements came from the 'fundamental concept of unity' (for arithmetic) and the concept of the 'dimensionless geometrical point' (for geometry) derived only from abstracted passively received stimuli. For Kant, the intuition of space formed the basis of geometry and the basis of arithmetic. Although their foundations for geometry and arithmetic were different, both regarded the disciplines as synthetic and constructive. The theorems of geometry were seen as rules to construct various figures and geometers were to discover necessary rules for their construction. Similarly, arithmetic was the search for rules governing the synthetic construction of numbers from other numbers (i.e. '5+7=12', two concepts synthesized into a single new concept). In both, the rules for construction were seen as necessary and not arbitrary, this being the restriction which led Kant to reject non-Euclidean geometry and consequently led many later philosophers to reject Kant's entire philosophy of mathematics.

At this time a more detailed inspection of the three main themes in Kant's mathematical philosophy, will be undertaken. These, again, are found mainly in the Critique of Pure Reason.

4. THE MAIN THEMES OF THE KANTIAN VIEW[6]

A. Mathematical Knowledge as Synthetic A Priori[7]

To be certain the statement "All mathematical judgements, without exception, are synthetic"(B14) is the cornerstone of Kant's mathematical philosophy. A priori knowledge requires strict necessity and strict universality neither of which can be found in a posteriori experience. For this reason Kant introduces the intuitions of space and time to account for the element of necessity imposed on a priori knowledge:

---

[6] Henceforth all references to Kant in the Critique will be denoted, in parentheses, by a letter (section) and a numeral (marginal numbering) referring to the translation given by Norman Kemp Smith (see references).

[7] In this part Kant's views from sections B1-B6 and B8-B17 of the Critique are presented.

> Mathematics gives us a shining example of how far, independently of experience, we can progress in a priori knowledge. It does, indeed, occupy itself with objects and with knowledge solely in so far as they allow of being exhibited in intuition. But this circumstance is easily overlooked, since this intuition can itself be given a priori, and is therefore hardly to be distinguished from a bare and pure concept. (B8)

Kant, in the Critique, then goes on to specify analytic or explicative judgements and synthetic or ampliative judgements in the way we saw earlier. Given his notions of these, we now see why it is so easy for Kant to rule out analytic a posteriori judgements:

> Since, in framing the judgement, I must not go outside my concept, there is no need to appeal to the testimony of experience in its support. (B12)

For synthetic judgements Kant cites two differing examples:

(1) All bodies are extended.
(2) All bodies are heavy.

For Kant, (1) can be seen as only tautologous and analytic while (2) supplies us with new information we did not presume with the concept of 'body'. In this respect he can now assert that all knowledge, in particular mathematical knowledge, if it is to give us new information, must be synthetic.

> Upon such synthetic, that is, ampliative principles all our a priori speculative knowledge must ultimately rest; analytic judgements are very important, and indeed necessary, but only for obtaining that clearness in the concepts which is requisite for such a sure and wide synthesis as will lead to a genuinely new addition to all previous knowledge. (B13-B14)

Building on this foundation Kant goes on to show explicitly how mathematical knowledge must be synthetic. In arithmetic '5+7=12' is a synthesis of two numbers into a new number. Although the equation could be misconstrued as analytic its syntheticity is evident when larger sums are performed and the answer is not necessarily very clear straight away. In geometry, a statement such as 'the shortest distance between two points is a straight line' must, similarly, be synthetic. In his own words:

> The concept of the shortest is wholly an addition, and cannot be derived, through any process of analysis, from the concept of a straight line. Intuition, therefore, must here be called in; only by its aid is the synthesis possible. (B16-B17)

Thinking the statement to be analytic is caused by the ambiguity of the terms, thus intuition (here, of space) must be added to join these otherwise disjoint concepts. For completeness Kant does note, however, that certain analytic statements in mathematics serve only as 'links in the chain of method' and not as principles. For example,

(1) a=a the whole is equal to itself
(2) (a+b)>a the whole is greater than its part

Even these concepts, though, are only admitted in mathematics because they are 'exhibited in intuition'.

Although the 'a priority' of mathematical assertions is often criticised[8], the synthetic nature of these judgements is usually what has caused contention. Toohey (1937, 2), in his article "Kant on the Propositions of Pure Mathematics", holds that Kant's views 'live or die' according to whether or not he shows that mathematical judgements are synthetic a priori, which Toohey belives he does not. To Toohey, there are two confusions in Kant's way of thinking. First, judging and uttering a proposition are distinct. To judge is to learn whereas in asserting a proposition one expresses what one has already learned from the act of judging. Secondly, making a proposition is not the same as repeating a judgement. To Toohey, it makes no sense to use terms like 'making the judgement again' because one never repeats the act of learning involved in judging unless one has suffered from something such as memory loss. Based on these confusions, Toohey states that Kant's account of an analytic judgement is inconsistent. If an analytic judgement is not a learning experience, in Kant's eyes, then it cannot be classified as a judgement at all. Attacking the problem from another angle Toohey notes that Kant, by an analytical judgement, implies that something which was confused has been made clear. Toohey observes that this is just another form of expanding ones' knowledge contrary to the way Kant has defined these statements. He exhorts that Kant even acknowledges this to some extent by stating: (Kant referring to 'All bodies are extended')

> For I do not require to go beyond the concept which I connect with 'body' in order to find [emphasis Toohey's] extension as bound up with it. (B11)

Thus Toohey's line of reasoning is an attempt to uproot Kantian ideas from the moment he (Kant) tries to show that mathematical propositions are exclusively synthetic a priori.

Considerably many other authors have had similar problems with the ambiguity Kant presents in the synthetic/analytic distinction (note that the question at issue here is 'does mathematics judge in Kant's way?' (if this way can be expressed coherently) and not whether it does so legitimately). Fischer (1866, 17) points out the ambiguity more closely by noting that the analyticity of a judgement cannot be established upon the fact that the two judgement concepts, A and B, are necessarily joined. In recognizing A and B as forming an analytic proposition Kant's phrasing implies "We must join this (B) to the concept (of A)" which could mean one of:

(1) B is a necessary part of concept A or
(2) B is a necessary addition or assertion about concept A

(1) is taken as analytic and (2) as synthetic by Fischer.

Paton (1936, I, 85) also alludes to this confusion. He notes that it appears that Kant's analytic judgements are mere definitions or are judgements stating the essence or part of the essence of a thing. He does note, however, "Kant's theory is not so simple as it looks, and the nature of analytic judgements is not altogether clear." He then adds later:

---

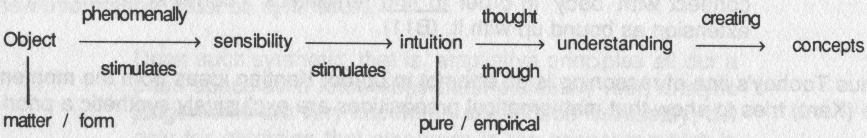
[8] For example, Fischer (1866) questions the necessity of mathematical judgements required for them to be a priori.

> Perhaps every judgement is essentially both analytic and synthetic, necessarily analysing a whole into its parts, and at the same time binding together the parts in a whole. Perhaps it is only in material operations, such as cutting wood and building houses, that analysis and synthesis, breaking up and putting together, can be separated from one another. (1936, I, 87)

Finally, one might ask, as Beck (1965, Ch. 4) has, whether Kant's synthetic judgements can be made analytic. In the 1760's when Kant first drew his analytic/synthetic distinction he (Kant) noted: "If one had the entire concept of which the notions of the subject and predicate are compars, synthetic judgements would change into analytic. It is a question of how much arbitrariness there is."[9] Apparently not much later Kant had decided that the dividing line between these two statement types was not variable or arbitrary. Beck attempts to discover why this was so and indicates the effect this has on Kant's overall philosophy. The extent to which Beck and others tackle the synthetic/analytic distinction is beyond the scope of this article and the reader, if interested, should refer elsewhere[10].

B. Intuition and Concepts - The Foundation of Knowledge[11]

How is synthetic a priori knowledge possible? To Kant, intuition relates to objects and is a mode of knowledge. The hierarchy of this perception is as follows:



The effect, in Kant's view, of an object upon the 'faculty of representation' is sensation and the intuition related to this object is empirical; the undetermined object of the perception comprising appearance. What is sensed in the appearance is matter (a posteriori). The manifold of appearance or form allows appearances to be relationally ordered a priori (because this ordering cannot be a posteriori part of the matter itself), apart from the actual sensation of matter. Kant then remarks, "The pure form of sensible intuitions in general, in which all the manifold of intuition is intuited in certain relations, must be found in the mind a priori. This pure form of sensibility may also itself be called pure intuition" (B35). Kant defines pure intuition (in the 'transcendental' sense), as opposed to empirical intuition, if nothing of this type of intuition belongs to the actual sensation of an object. Thus if we, in the representation of a body, subtract away the facets of empirical intuition in the understanding (substance, force, divisibility, etc.) and sensation (colour, penetrability, etc.) we are left with extension and figure comprising pure intuition "which, even without any actual object of the senses or of sensation, exists in the mind a priori as a mere form of sensibility". (B35)

---

[9] Kant, Reflection 3928 (Academy edn., XVII, 350)

[10] For example see Hyslop (1903, 331-351)

[11] In this part Kant's views from sections A101-A103, B64-B66, B146-B147 and B297-B299 of the Critique are presented.

Kant further distinguishes between thinking and knowing an object. As the schematic shows, knowledge requires a concept (the category) through which the object is thought and intuition through which the object is given. The forms of sensible intuitions are either pure (space/time, extension/figure) or empirical (that which can be represented as 'actual' in space and time). He emphasizes that pure intuition is not alone a source of knowledge of phenomenon.

> Through the determination of pure intuition we can acquire a priori knowledge of objects, as in mathematics, but only in regard to their form, as appearances; whether there can be things which must be intuited in this form is still left undecided. (B147)
>
> Consequently, the pure concepts of understanding even when they are applied to a priori intuitions, as in mathematics, yield knowledge only in so far as these intuitions - and therefore indirectly by their means the pure concepts also - can be applied to empirical intuitions. (B147)

Concerning the meaning of concepts in mathematics Kant cites the example of the concept of the intuition of space which, although a priori, is meaningless unless its meaning is represented in appearances or empirical objects. "Otherwise the concept would, as we say, be without sense, that is, without meaning" (B298). Another example he uses is from geometry where a 'figure' is constructed to meet this demand. Although the figure is produced a priori it serves as a public empirical appearance to the senses. Similarly in the concept of magnitude, although a priori, it is supported by fingers, beads of the abacus, or strokes or points placed before the eyes.

> The concept itself is always a priori in origin and so likewise are the synthetic principles or formulas derived from such concepts; but their employment and their relation to their professed objects can in the end be sought nowhere but in experience, of whose possibility they contain the formal conditions. (B299)

Space and Time - Subjective Facets of Pure Intuition

Kant's philosophy of space and time, which underly geometry and arithmetic, is firmly rooted in the above explicated theory of perception. Kant begins by assuming space and time to be objective. Since geometric propositions are synthetic a priori and are known with certainty how have they come to be necessary and universally valid truths in the first place? The only way, according to Kant, is through concepts or intuitions given either a priori or a posteriori. If given a posteriori, through empirical intuitions, neither the concepts nor the intuitions can yield non-empirical synthetic propositions and for this reason geometric propositions of this kind can never possess the necessity and absolute universality which are supposedly characteristic of them. On the other hand, if these concepts have been given a priori, then from mere concepts (and no empirical basis) only analytic, not synthetic knowledge is obtainable. For example, the proposition 'two straight lines cannot enclose a space, and with them alone no figure is possible' cannot be derived from the concepts of the straight line and the number two. Thus one must obtain an object through intuition to form a synthetic statement. It cannot be given in empirical intuition in which no universal statements are possible. From this point Kant advances in the following way:

> If there did not exist in you a power of a priori intuition; and if that subjective condition were not also at the same time, as regards its form, the universal a priori condition under which alone the object of this outer intuition is itself possible; if the object (the triangle) were something in itself, apart from any relation to you, the subject, how could you say that what necessarily exist in you as subjective conditions for the construction of a triangle, must of necessity belong to the triangle itself? You could not then add anything new (the figure) to your concepts (of three lines) as something which must necessarily be met with in the object, since this object is (on that view) given antecedently to your knowledge, and not by means of it. If, therefore, space (and the same is true of time) were not merely a form of your intuition, containing conditions a priori, under which alone things can be outer objects to you, and without which subjective conditions outer objects are in themselves nothing, you could not in regard to outer objects determine anything whatsoever in an a priori and synthetic manner. (B66)

Thus Kant concludes the certainty that space and time must be subjective conditions of our intuition[12] and in relation to these conditions all objects are mere appearances not perceived as things-in-themselves existing in an objective space or time.

Tackling the problem of objections to Kant's views of space and time is a difficult task. Certainly his views have had significant influence on the philosophy of Euclidean geometry as well as influencing the cosmology of Einstein and Minkowski in their denial of objective space-time. However, subsequent philosophy to Kant has usually been very nonaccepting of his ideas. Some of these difficulties are briefly sketched here.

Fischer (1866) notes that Kant assumes necessity to be a proof that the concept of space is a priori. He retorts that contrary to this are schools of thought, psychologically based, which claim to exhibit cases of necessary beliefs generated, but found to be false. As well, Fischer questions Kant's acceptance of the independence of space and time noting J.S.Mill's approach which attempts to deduce space from time (but Fischer notes the conceptual difficulties this also entails). With regards to intuition and its role in mathematics Stebbing (see Paton 1936, I, 156) remarks, "Mathematics can be exhibited as a completely logical structure, so that no element of intuition enters into a mathematical proof" reflecting much of the modern sentiment towards the Kantian outlook. In Kant's defence, Paton remarks that (logical) system interpretation involves finding objects for undefined concepts - these objects, ultimately, must be perceived somehow in space and time. On the other hand, Paton (1936, I, 146-152) criticizes Kant"s proof of the universality of space and time. Evidently much of this anti-Kantian sentiment has stemmed from the difficulty one has in accepting the subjectivity of space and time.

---

[12] Kant compares himself with Copernicus, and his discovery of the rotation of the earth, in making this revolutionary insight (preface to the second edition of the Critique p. XXXIX)

Much controversy has been made of Kant's Euclidean conception of geometry as well. Ernst Cassirer (1953), in his book (a compilation of two works), Substance and Function - Einstein's Theory of Relativity, quarrels with the a priori nature of Kant's axioms of geometry. He holds that the a priori freedom of the mathematical concept and its separation from all sensuous appearance is 'not properly articulated' in the Critique. In his more 'pragmatic' approach, Cassirer sees modern mathematics as choosing the geometry which best fits the physical reality of the situation. "It is reasoned that, as all the systems are equally valid in logical structure, we need a principle that guides us in their application. This principle can be sought only in reality,..." (1953, 106). Effectively, the pure a priori nature of the Kantian axioms must be questioned, this being the key problem in reconciling Kant to modern non-Euclidean trends. The range of objections to Kant's framework of perception also delve deeply into the philosophy of science and reality as well as mathematics. However this, again, would take us beyond our desired scope.

C. Mathematics as a Constructive Enterprise[13]

> Mathematics presents the most splendid example of the successful extension of pure reason, without the help of experience. (B740)

Part of Kant's motivation to uncover the mechanism of mathematics was to compare and contrast this mechanism with philosophy. Mathematical knowledge is gained by reason from concept construction as opposed to philosophy which is knowledge gained through reason from concepts themselves. In mathematics, for the construction of a concept, pure intuition is needed. Thus representation of, for example, a triangle, can be in the imagination (pure intuition) or on paper (empirical intuition), in both cases completely a priori; not having borrowed the pattern from any experience. In the act of this construction on paper its universality remains unchanged not altering the concept of 'triangle' with differing dimensions. Thus, due to the nature of mathematics, it considers the universal in the particular as opposed to philosophy which considers the particular only in the universal. This is the formal distinction between the two, Kant draws.

Kant then goes deeper with his contrast between philosophers' and mathematicians' ways of coming to knowledge. Kant proposes that, given the concept of a triangle, a philosopher would not be able to find the relation between the sum of its angles and a right angle. "However long he meditates on this concept, he will never produce anything new. He can analyse and clarify the concept of a straight line or of an angle or of the number three, but he can never arrive at any properties not already contained in these concepts" (B744). In contrast, the geometrician would solve the problem differently. His reasoning would begin at constructing the concept of a triangle with pure intuition, optionally exhibiting the construction empirically. Through a chain of inferences during the construction, the geometer would then arrive at a fully evident universally valid solution of the problem.

As in the construction of magnitudes (quanta) for geometry, mathematics also constructs magnitudes (quantitas) in arithmetic. In arithmetic, a magnitude is abstracted completely from the properties of the object that it is to be thought in terms of. A notation of magnitude construction (numbers) is then adopted. Again, Kant explains himself best on this point:

---

[13] Kant's views from sections B740-B747 of the Critique.

> Once it (mathematics) has adopted a notation for the general concept of magnitudes so far as their different relations are concerned, it exhibits in intuition, in accordance with certain universal rules, all the various operations through which the magnitudes are produced and modified. When, for instance, one magnitude is to be divided by another, their symbols are placed together, in accordance with the sign for division, and similarly in the other processes; and thus in algebra by means of a symbolic construction, just as in geometry by means of an ostensive construction (the geometrical construction of the objects themselves), we succeed in arriving at results which discursive knowledge could never have reached by means of mere concepts. (B745)

This is the method by which Kant proposes that synthetic a priori mathematical judgements can be formulated. That is, by the construction of the concepts through pure intuition, in contrast to the route of the philosopher who has open to him only the synthesis of concepts already existing and not constructed. Kant then goes deeper into the logic of mathematical reasoning but this is beyond the compass of explicating the basic philosophy of his constructive mathematical views.

It must be mentioned that Kant's view of mathematics as constructive has fallen under just as much criticism as his other views explicated earlier. These center on the role of intuition in the construction of concepts and the construction of algebraic and geometric concepts. And again, much criticism has also focussed upon Kant's apparent requirement that geometric concepts be constructed as Euclidean. However it would not seem incredible, in my view, to attempt an adaptation of Kantian views to modern geometries in preference to rejecting all of the relevance of Kant's views to current foundational research. Today's constructivists and intuitionists have shown this type of restraint, modern influences, of which they are an example, being the topic of the final brief and closing section of this survey.

5. MODERN KANTIAN INFLUENCES

Along with profoundly influencing the positivism of Comte and later ultimately of the Vienna Circle[14], Kantian influences persist to the present day. Indirectly, Kant has influenced the formalism of Hilbert[15] which arose out of the problem of reference in geometry. Geometry refers to space, as Kant firmly believed in his time, however the arisal of non-Euclidean geometries caused this idea to break down for some philosophers such as Hilbert. Hilbert repostulated, then, that geometry did not really 'refer', in the strict sense, but was more of a formal game of which there were rules to follow. Another major Kantian influence on Hilbert came from the Critique of Practical Reason in which Kant suggested that, due to the unattainability (through knowledge) of noumenal concepts such as God, one should live 'as if' the world were created by God. Hilbert latched on to this concept and attempted to talk in an 'as if' manner about the existence of actual infinity. As Tucker (1985, 32) puts it, "The Kantian approach tofoundations is an attempt to justify the appeal to actual infinity by interpreting it as no more than an 'as if' view. It is the view that mathematicians must continue to talk as if, and only as if, there were actual infinities, but without really meaning that there are such things."

---

[14] A circle of thinkers (the 'logical positivists') including many eminent scientists and mathematicians (Godel, Carnap, etc.) which formed in Vienna in the 1920's. See Neurath and Cohen (1973).

[15] Hilbert, in his introduction to Grundlagen du Geometrie, quotes from Kant's Critique: "Thus all human knowledge begins with intuition, goes from there to concepts, and ends with ideas."

Another major influence of Kant on modern mathematical thought is on the current intuitionist movement. The intuitionist, borrowing the 'intuition' concept on Brouwer's lead, looks upon the most rudimentary intuition (Kant's inner sense) as the awareness of successive mental states. This causes, for them, through the intuition of time (as 'successive'), the possibility of constructing a linear series. This is the implicitness of the natural numbers in the stream of our consciousness; thus arithmetic is conceived as growing directly out of our primary awareness. Intuitionists insist that conceptual thinking in mathematics is not competent to put anything into mathematics that cannot already be found in the intuition itself. Also, mathematical existence is coincident with constructibility in this philosophy. Thus, as a philosophy, intuitionism bears some resemblance to Kant's ideas.

6. CONCLUSION

From what has been previously foreclosed, any study of philosophy of mathematics can certainly not be complete without a look into the critical philossophy of Immanuel Kant. The hope is that it can at least be seen from this survey that Kantian ideas, although obscure and often controversial, contain substance; a substance strong enough to have reached into the twentieth century and affect today's foundational questions.

REFERENCES

Baum, R.J. Philosophy and Mathematics, from Plato to the Present. San Francisco: Freeman, 1973.

Beck, L.W. Studies in the Philosophy of Kant. Indianapolis: Bobbs-Merrill, 1965

Cassirer, Ernst. Kant's Life and Thought, trans. James Haden. London & New Haven: Yale University Press, 1981.

-----, Substance and Function - Einstein's Theory of Relativity, trans. W.C.Swabey and M.C.Swabey. New York: Dover, 1953.

Fischer, Kuno. A Commentary on Kant's Critick of the Pure Reason, trans. Kuno Fischer. London: Longmans, 1866.

V. Freytag gen. Loringhoff, Dr. Bruno Baron. Philosophy Problems of Mathematics, trans. Amethe Countess von Zeppelin. New York: Philosophical Library, 1951.

Hintikka, J. "Kant on the Mathematical Method." The Monist, vol. LI, 1967 (July).

Hyslop, J.H. "Kant's Treatment of Analytic and Synthetic Judgement", The Monist, vol. XIII, 1903.

Kant, Immanuel. Critique of Pure Reason, trans. Norman Kemp Smith, 2nd edn. London: Macmillan and Canada: Macmillan, 1933.

Kneebone, G.T. Mathematical Logic and the Foundations of Mathematics - An Introductory Survey. London, Princeton, Toronto: D. van Nostrand, 1963.

Lindsay, A.D. Kant (London: Benn, 1934), chapter II, "Kant's Pre-Critical Writings and his relations to his predecessors."

Maziarz, Edward A. The Philosophy of Mathematics. New York: Philosophical Library, Inc., 1950.

Neurath, M. and Cohen, R.S., eds. Otto Neurath : Empiricism and Sociology. Dordrecht: Reidel, 1973.

Parsons, C. "Kant's Philosophy of Arithmetic" in Morgenbesser, S., Suppes, P., and White, M., Philosophy, Science and Method. New York: St. Martin's Press, 1969.

Paton, H.J. Kant's Metaphysic of Experience, 2 vols. London: Allen, and New York: Macmillan, 1936.

Toohey, J.J. "Kant on the Propositions of Pure Mathematics." The New Scholasticism, vol. II, 1937.

Tucker, John. "Constructivity, Consistency and Natural Languages." Five Papers on the Foundations of Mathematics. (University of Waterloo, 1985)

Walsh, W.H. Kant's Criticism of Metaphysics. Edinburgh: Edinburgh University Press.

Wilder, R.L. Introduction to the Foundations of Mathematics, 2nd edn., New York, London, Sydney: John Wiley & Sons, 1965.

PROBLEMS DRIVE

Chris Kilgour (St. John's College) and **Matthew Richards** (Trinity College)

Restrict yourself to five minutes per question!

1. Which is bigger,

   $\sqrt{(5/12)} + 1/\sqrt{5}$ or $1/\sqrt{3} + \sqrt{(2/7)}$?

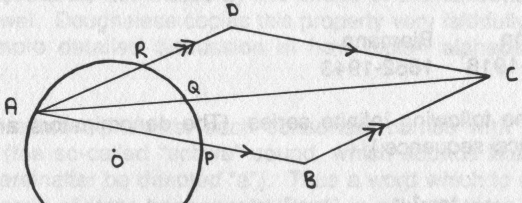
2. What is the area of the largest quadrilateral with sides of length 1, 2, 3, and 4?

3. Five Part II logicians are discussing an alleged vote-rigging scandal at elections to the Archimedeans' committee. They make statements as follows :

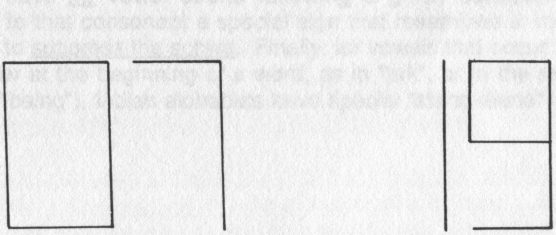
   Arthur: "I was with Edward all the time. Charles and Gordon were involved."
   Charles: "I am innocent. Only one of Edward and Gordon was involved."
   Edward: "Charles and Gordon are both innocent."
   Gordon: "Ingrid is innocent, but Edward is guilty."
   Ingrid: "Two people were involved, neither of which was Gordon or myself."

   Assuming they were not all involved, and the innocent tell the truth while the guilty lie or not as they choose, who was involved?

4. In this diagram, if AB=7, AC=8, AD=4, AP=4, and AR=3, what is AQ?



5. A digital alarm clock uses the familiar seven-segment numbers to display the time in hours and minutes in the 24-hour clock, for example as in the picture below. Which of the 28 segments is lit for the longest total time in a day, and which for the least?



6. Give the next number in each of the following sequences:

   (i) 2, 3, 10, 26, 72, ...
   (ii) 2, 3, 10, 45, 236, ...
   (iii) 2, 3, 10, 12, 13, ...
   (iv) 2, 3, 10, 12, 21, ...

7. Given an equilateral triangle ABC with sides of length 1, what is the length of the shortest curve from an interior point of AB to an interior point of AC that halves the area of the triangle?

8. The galaxy of Qarch contains five peculiar planets which are in the shapes of the five Platonic solids with edges of length 100 Archimedean miles. The interiors of all the faces are covered by sea, leaving only the edges to walk on. An astronaut is landed at a vertex of each planet with instructions to explore all accessible parts of it. Which astronaut has to walk the furthest, and which the least, and how far in each case? (They do not have to return to their starting points.)

9. Given a triangle ABC of perimeter p, and a point X outside ΔABC, show how to construct a line through X which cuts off a triangle of perimeter p/2 from ΔABC.

10. Match the following mathematicians with their correct dates:

Cantor	Euler	Fermat	Galois	Gauss	Hilbert
1601-1665	1643-1727	1707-1783	1736-1813	1777-1855	1811-1832

Lagrange	Newton	Riemann
1826-1866	1845-1918	1862-1943

11. Find the sum of the following infinite series. (The denominators are alternate terms of the Fibonacci sequence.)

$$\tan^{-1} 1 + \tan^{-1}\frac{1}{2} + \tan^{-1}\frac{1}{5} + \tan^{-1}\frac{1}{13} + \tan^{-1}\frac{1}{34} + \ldots$$

12. Find the smallest integer greater than

$$1 + \frac{1}{\sqrt{2}} + \frac{1}{\sqrt{3}} + \ldots + \frac{1}{\sqrt{(1988^2 - 1)}} + \frac{1}{1988}$$

INTRODUCTION TO DOUGHALESE

Douglas R. Hofstadter

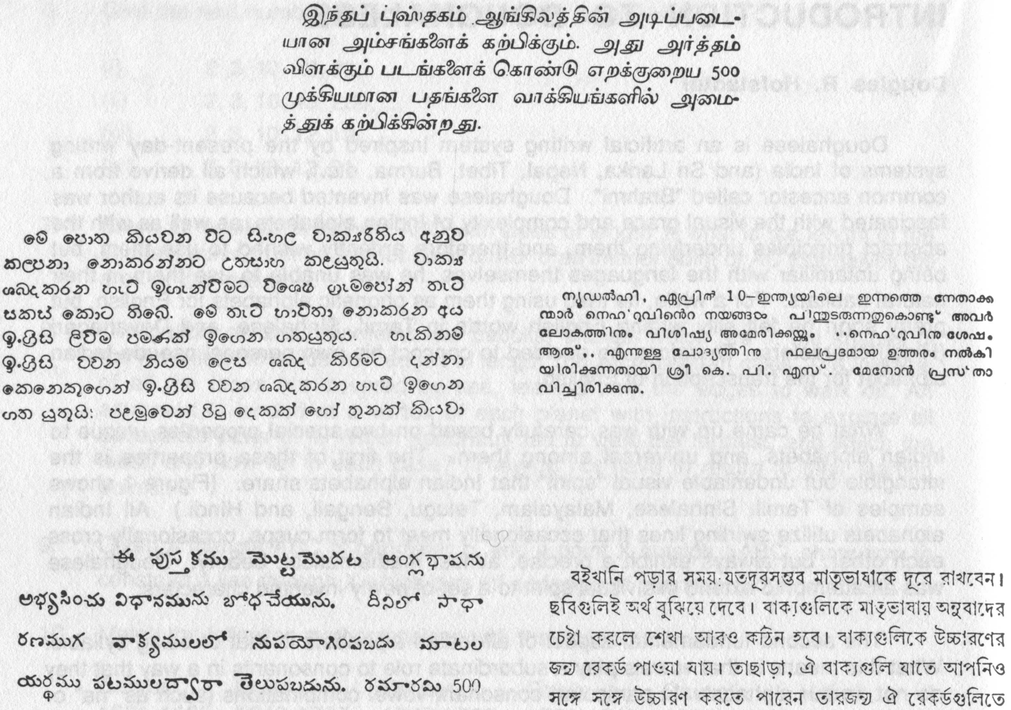
Doughalese is an artificial writing system inspired by the present-day writing systems of India (and Sri Lanka, Nepal, Tibet, Burma, etc.), which all derive from a common ancestor called "Brahmi". Doughalese was invented because its author was fascinated with the visual grace and complexity of Indian alphabets, as well as with the abstract principles underlying them, and therefore ardently wished to use them, but being unfamiliar with the languages themselves, he was unable to use them in their natural habitats. For a while, he tried using them as phonetic alphabets for English, but pretty soon he felt silly writing English words in Tamil, Sinhalese, and Devanagari (Hindi) characters. Therefore, he decided to concoct his own personal pseudo-Indian alphabet for the transcription of English.

What he came up with was carefully based on two special properties unique to Indian alphabets, and universal among them. The first of these properties is the intangible but undeniable visual "spirit" that Indian alphabets share. (Figure 1 shows samples of Tamil, Sinhalese, Malayalam, Telugu, Bengali, and Hindi.) All Indian alphabets utilize swirling lines that occasionally meet to form cusps, occasionally cross each other, but always exhibit a precise, almost mathematical, beauty. Doughalese was an attempt to extend this visual spirit to a set of newly invented characters.

The second fundamental aspect of all Indian alphabets is that of being syllabic. What this means is that vowels play a subordinate role to consonants in a way that they do not, in our alphabet. In particular, consonant-vowel combinations (such as "na" or "gu") are fused units dominated by the shape of the consonant, and modified by the following vowel. Doughalese copies this property very faithfully. In order to explain this aspect, a more detailed discussion of how Indian alphabets represent sounds is required.

In all Indian alphabets, each consonant comes with a default neutral vowel following it (the so-called "schwa" sound, which sounds like the "a" in "dollar", and which will hereinafter be denoted "a"). Thus a word which to a European eye appears to be composed of three consonants ("krn") is actually three syllables whose vowels are all schwas ("karana"). The schwa vowel is easily overridden by attaching special subordinate-vowel markers that stand for "**a**" ("long a" as in "father"), "e" (as in "met"), "i" (as in "machine"), "o" (as in "hope"), or "u" (as in "flute"). Actually, in many Indian languages there are considerably more vowel distinctions than this, but since those are the only ones in Doughalese, we will not describe those distinctions.

In order to have no vowel sound following a given consonant, most Indian languages attach to that consonant a special sign that resembles a vowel marker, but that serves solely to suppress the schwa. Finally, for vowels that occur at the beginning of a syllable (either at the beginning of a word, as in "ark", or in the middle of a word, such as the "i" in "being"), Indian alphabets have special "stand-alone" vowel symbols.



இந்தப் புஸ்தகம் ஆங்கிலத்தின் அடிப்படையான அம்சங்களைக் கற்பிக்கும். அது அர்த்தம் விளக்கும் படங்களைக் கொண்டு எறக்குறைய 500 முக்கியமான பதங்களை வாக்கியங்களில் அமைத்துக் கற்பிக்கின்றது.

මේ පොත කියවන විට සිංහල වාක්‍යරීතිය අනුව කල්පනා නොකරන්නට උත්සාහ කළයුතුයි. වාක්‍ය ශබ්දකරන හැටි ඉගැන්වීමට විශේෂ ග්‍රැමපෝන් තැටි සකස් කොට තිබේ. මේ තැටි භාවිතා නොකරන අය ඉංග්‍රීසි ලිවීම පමණක් ඉගෙන ගතයුතුය. හැකිනම් ඉංග්‍රීසි වචන නියම ලෙස ශබ්ද කිරීමට දන්නා කෙනෙකුගෙන් ඉංග්‍රීසි වචන ශබ්දකරන හැටි ඉගෙන ගත යුතුයි: පළමුවෙන් පිටු දෙකක් හෝ තුනක් කියවා

ന്യൂഡൽഹി, ഏപ്രില് 10—ഇന്ത്യയിലെ ഇന്നത്തെ നേതാക്കന്മാർ നെഹ്റുവിന്റെ നയങ്ങളും പിന്തുടരുന്നതുകൊണ്ട് അവർ ലോകത്തിനും വിശിഷ്യ അമേരിക്കയ്ക്കും ''നെഹ്റുവിനു ശേഷം ആര്''. എന്നുള്ള ചോദ്യത്തിനു ഫലപ്രദമായ ഉത്തരം നൽകിയിരിക്കുന്നതായി ശ്രീ കെ. പി. എസ്. മേനോൻ പ്രസ്താവിച്ചിരിക്കുന്നു.

ఈ పుస్తకము మొట్టమొదట ఆంగ్లభాషను అభ్యసించు విధానమును బోధచేయును. దీనిలో సాధారణముగ వాక్యములలో నుపయోగింపబడు మాటల యర్థము పటములద్వారా తెలుపబడును. రమారమి 500 ముఖ్యమగు మాటల భావము ఆవిధముగ నీపుస్తకమున అనుసరింపబడినది.

বইখানি পড়ার সময় যতদূরসম্ভব মাতৃভাষাকে দূরে রাখবেন। ছবিগুলিই অর্থ বুঝিয়ে দেবে। বাক্যগুলিকে মাতৃভাষায় অনুবাদের চেষ্টা করলে শেখা আরও কঠিন হবে। বাক্যগুলিকে উচ্চারণের জন্য রেকর্ড পাওয়া যায়। তাছাড়া, ঐ বাক্যগুলি যাতে আপনিও সঙ্গে সঙ্গে উচ্চারণ করতে পারেন তারজন্য ঐ রেকর্ডগুলিতে সময়ের ব্যবস্থা আছে। যদি এই রেকর্ড না থাকে তবে বইয়ে

चीन के सात प्रतिनिधि चीनी साम्यवादी पार्टी के सेक्रेटरी-जनरल तेंग ह्स्याओ-पिंग के नेतृत्व में मास्को गये । शुक्रवार ५ जुलाई को उन्होंने रूसी प्रतिनिधियों से वार्तालाप शुरू किया । सुस्लोव रूसी प्रतिनिधि-दल के नेता थे । वार्ताएं अत्यंत गुप्त रखी गयीं । किंतु यह बात छिपी नहीं रही कि रूस और चीन के बीच के मतभेदों की खाई भरने की जगह और चौड़ी हो गयी । अंत में ‘प्रवदा’ ने चीन का १४ जून का पत्र एवं उसका विस्तृत उत्तर छापा और चीन पर पिछड़े देशों को गुमराह करने और स्टालिन की दमन-नीति का समर्थन करने का आरोप लगाया । वार्ताएं लंगड़ा-लगड़ाकर चल रही हैं । दोनों देशों के रेडियो और अखबार आग उगलने लगे हैं ।

Figure 1.

Doughalese, following these models, assigns to each consonant the default schwa sound, and has subordinate-vowel markers, stand-alone vowel markers, and a schwa-suppression sign. In Doughalese, as in many Indian languages, most of the vowel markers (and the schwa suppressor) are written above the consonant. There are, however, exceptions. The sign for "**a**" ("long a") follows the consonant (i.e., goes to the right of it), and the sign for "u" goes underneath the consonant, mirroring the standard practices in most Indian alphabets.

Although Doughalese was devised to transcribe English, it nonetheless does not make all the phonemic distinctions of English, since it is based on the Indian languages, whose set of phonemes is quite different from that of English. As far as consonants are concerned, Doughalese has only one symbol for the distinct sounds "b" and "p", only one symbol for the distinct sounds "t" and "th" (both voiced and unvoiced, as in "breathe" and "breath"), and only one symbol for the distinct sounds "s" and "z". Doughalese has a special symbol for the English "sh" sound, which also covers "zh" (the sound of "s" in "measure"), but none for "ch", "j", or "x"; these sounds are, instead, represented as compound consonants. Thus, "ch" is represented as "t" followed by "sh" (think of "boatshow"), "j" as "d" followed by "sh" ("midship"), and "x" as "ks" ("thanks").

English has a much richer vowel system than Doughalese does. For instance, the three different vowel sounds in "bat", "bet", and "bait" are all represented in Doughalese by the same sign -- the "e" vowel. Thus the Doughalese syllable consisting of the "p-b" sound combined with the "e" vowel, followed by the "t-th" sound combined with the schwa-suppressor, is highly ambiguous -- it could stand for a plethora of English words, including "bat", "bet", "bait", "bath" (as pronounced in America), "Beth", "bathe", "pat", "pet", "pate", and "path" (also as pronounced in America), as well as the non-words "peth" and "pathe". Similarly, there is no distinction in Doughalese between the vowels in "bit" and "beat" -- they are both "i". And the vowels in "book" and "boot" are both represented by the "u" sign. Furthermore, the schwa stands not only for the sound of "a" in "dollar", but also for the sound of "u" in "bug". Finally, to represent the sound "ay" (as in "aye", "eye", or "fine"), Doughalese uses an "**a**" sign followed by an "i" sign, and to represent the "ow" sound (as in "bowwow"), it uses an "**a**" sign followed by a "u"-sign.

To give a sense of how English is transcribed into Doughalese, this very sentence will now be converted into its Doughalese equivalent written in roman letters, with the schwa-suppressor indicated by a dagger superscript (thus any consonant without a vowel or schwa-suppressor after it has an implicit schwa-sound following it).

> Tu giv† a sen†s† av† h**a**u ing†lish† is† t†ren†s†k†r**a**ib†d† in†tu dglis†, tis† veri sen†tn†s† vil† n**a**u bi† cn†vr†td† in†tu it†s† dglis† ikuivln†t† ritn† in† romn† letr†s†, vit† t shu**a**-sp†resr† in†diketd† b**a**i a degr† supr†s†k†rip†t† (ts† eni k**a**n†snn†t† vit**a**ut† a v**a**ual† or† shu**a**-sp†resr† ef†tr† it† hes† en† im†p†list† shu**a**-s**a**un†d† faloing† it†).

The underlines under "sh" and "ng" are to remind you that those two-letter combinations are represented by single symbols in Doughalese.

One final curious property of Doughalese is that it has two distinct symbols for "n", one to be used only at the beginnings of syllables (as in "nag"), the other (which coincides with the "ng" symbol) only at the ends of syllables (as in "kin" or "king"). This also reflects a distinction, present in most Indian languages, between different types of "n" sounds.

The chart of Doughalese symbols (Figure 2) is based on the order in which Indian alphabets are traditionally presented (and a quite logical order it is, too):

a	i	u	e	o	<--	(short vowels)
a	**i**	**u**	**e**	**o**	<--	(long vowels)
k	kh	g	gh	ng		
ch	chh	j	jh	ñ		
t	**th**	**d**	**dh**	**n**	<--	(palatal consonants)
t	th	d	dh	n	<--	(dental consonants)
p	ph	b	bh	m		
y	r	l	v			
sh	s	h	f			

Doughalese makes no distinction between unaspirated and aspirated consonants ("k" and "kh", for instance, or "b" and "bh"). This means that the first and second columns of consonants are collapsed into a single column in Doughalese; similarly, the third and fourth columns collapse into a single column. Thus in the Doughalese consonant chart, there are only three columns, instead of five.

Also, Doughalese makes no distinction between palatal and dental consonants, so the third and fourth rows of consonants are collapsed into a single row. And finally, Doughalese has no "ch" row, so that in the Doughalese consonant chart, there are only three rows, instead of five.

Figure 3 is an extended passage in Doughalese. Readers may enjoy trying to read it without help; however, for those who may not have the patience to do so, here is a line-by-line transcription into roman characters:

Tis† is† a sem†pl†
av† t siudoin†dian†
raiting† sis†tm† in†ven†td†
bai Dg†ls† Haf†s†tetr†
in† t mid†sik†s†tis†, ven†
hi vs† a met† med†shr†
et† S†ten†for†d† Yunivr†siti.

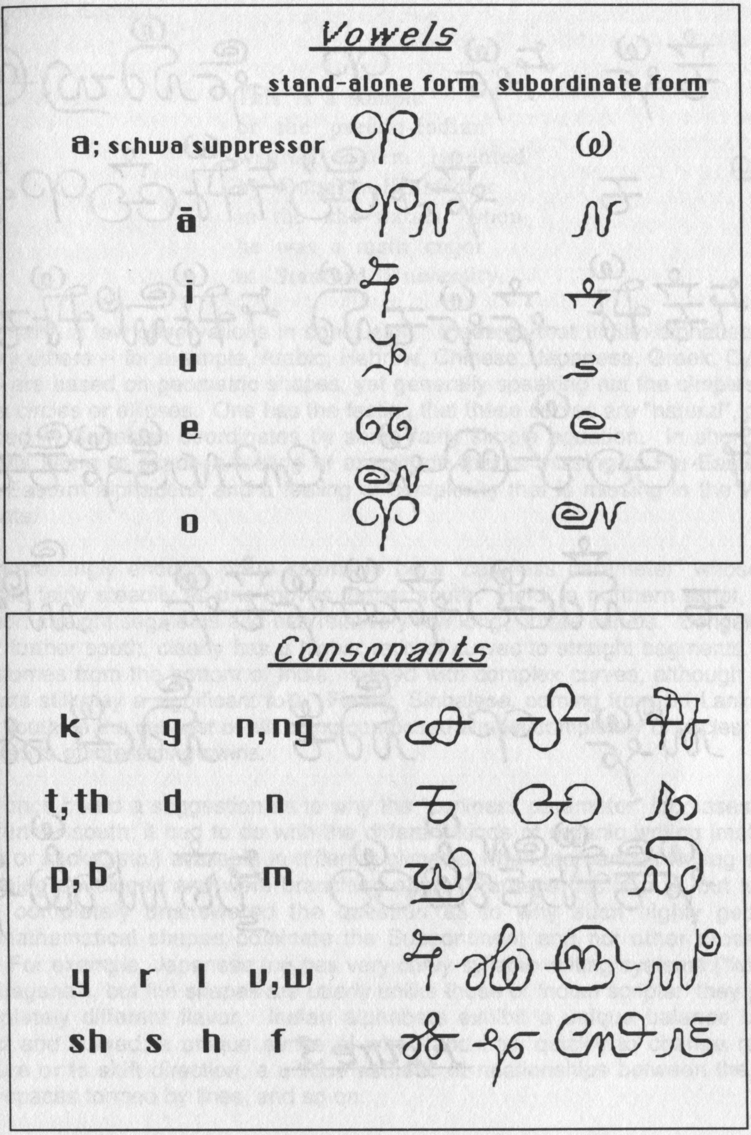

Vowels
stand-alone form
subordinate form
a; schwa suppressor
ā
i
u
e
o
Consonants
k
g
n, ng
t, th
d
n
p, b
m
y
r
l
v, w
sh
s
h
f

Figure 2.

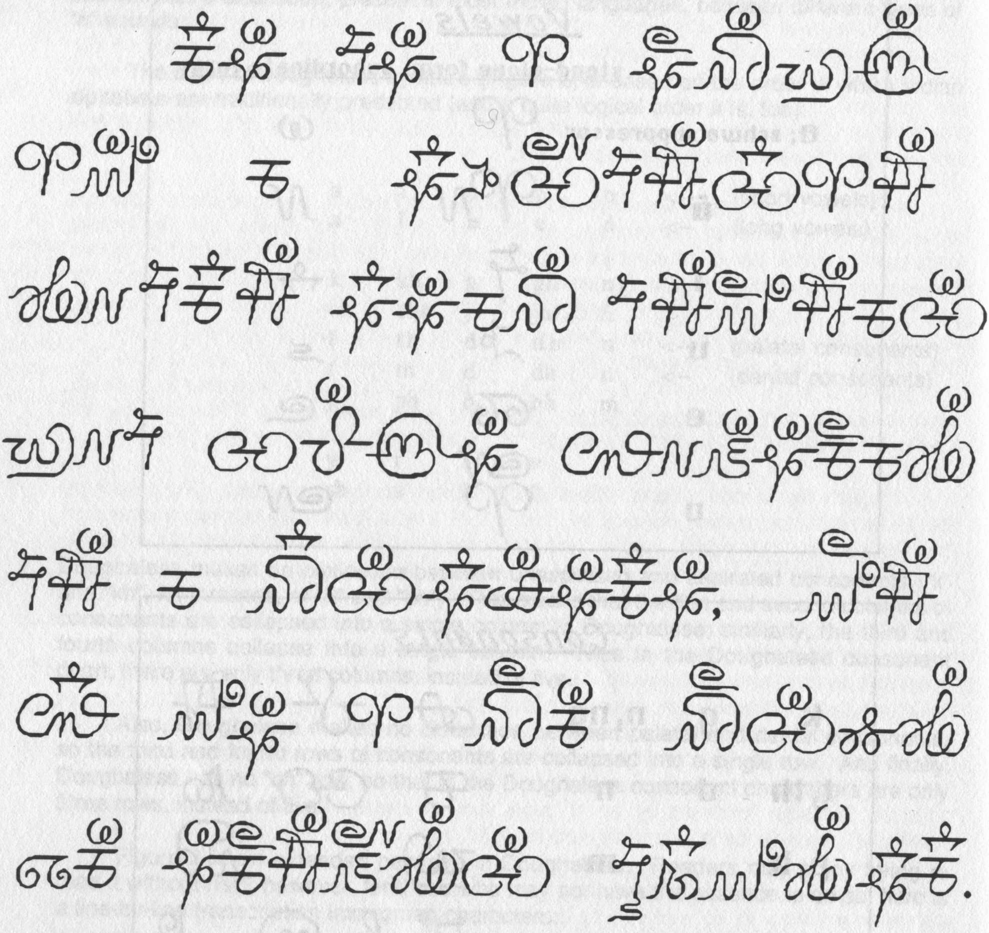


Figure 3.

and in normal English:

This is a sample
of the pseudo-Indian
writing system invented
by Douglas Hofstadter
in the mid-sixties, when
he was a math major
at Stanford University.

Finally, a few observations in conclusion. It seems that Indian alphabets, more than any others -- for example, Arabic, Hebrew, Chinese, Japanese, Greek, Cyrillic, or Latin -- are based on geometric shapes, yet generally speaking not the simplest ones, such as circles or ellipses. One has the feeling that these curves are "natural", perhaps described in Cartesian coordinates by some fairly simple equation. In short, Indian alphabets seem to exude a feeling of exactitude that is missing in Far-Eastern and Middle-Eastern alphabets, and a feeling of complexity that is missing in the Western alphabets.

Interestingly enough, there seems to be a "curliness parameter" whose value increases fairly steadily as one moves further south. Hindi, a northern script, is filled with short straight segments and has relatively few long, ornate curves. Bengali, which is a bit further south, clearly has a higher ratio of curves to straight segments. Tamil, which comes from the bottom of India, is filled with complex curves, although straight segments still play a significant role. Finally, Sinhalese, coming from Sri Lanka, even further south, is the curviest of all, being composed almost completely of circles, spirals, and all sorts of interacting swirls.

I once heard a suggestion as to why the "curliness parameter" increases as you move further south; it had to do with the different kinds of organic writing implements (leaves or sticks, etc.) available in differing climates when the various writing systems were being developed and were branching apart. Perhaps this is true, but for me it leaves completely unanswered the question as to why such highly geometric, quasi-mathematical shapes dominate the Subcontinent and not other areas of the world. For example, Japanese too has very curvy syllabic writing systems ("katakana" and "hiragana"), but the shapes are utterly unlike those of Indian scripts: they possess a completely different flavor. Indian alphabets exhibit a unique balance between straight and curved, a unique sense of when and how quickly to change radius of curvature or to shift direction, a unique esthetic of relationships between the various hollow spaces formed by lines, and so on.

Surely, something in the collective genius of the Indian peoples -- not just their writing implements and their history -- determined these beautiful spirits behind the letters, and I would love to see that discussed (and not just in a mystical way, relating Tamil letters, as I once saw done, to the forms of embryos in the womb!). I would like to see a study that relates the geometry of Indian letterforms to other aspects of Indian civilization, including Indian music, mathematics, etc. But I would guess that no such study has ever been done.

When I was first falling in love with the Indian scripts, I experienced very strongly the mysterious fact that pure shape can, in fact, be much more than pure shape: it can contain and convey ideas. This notion -- that ideas can lurk hidden in complex swirls -- could be called the "semiotic mystery". We get so used to it that we are almost never amazed by it; only intense contact with other writing systems tends to awaken us from our usual blasé state on this score. For a while, my immersion in Indian alphabets, culminating in my development of and practice with Doughalese, allowed me to inhale very freshly and powerfully the exotic scent of the semiotic mystery, but alas, as in any love affair, the power and mystery gradually faded. That special weird feeling is no longer evoked when I let my eye roll over a page printed in Malayalam, Tamil, or Sinhalese; nonetheless, I hope that this article may perhaps inspire others to indulge in the deep joy of Indian alphabets, and to experience the semiotic mystery that they once so richly exuded for me.

AN INTRODUCTION TO WARGAMING

Paul Fernandez

Chess is a remarkable game. Thirty-two pieces, sixty-four squares, traditions dating back centuries and new ideas discovered every month. However, I have to admit to finding chess somewhat unsatisfying. The very thing that makes it so attractive to many people - the abstraction of strategic and tactical principles - to me makes chess formal and stagnant. This is not to say I dislike it, but neither do I take great pleasure in it.

It was after losing my captaincy of the school chess team that I discovered a book in the public library about board wargaming. I read it from cover to cover, and then - unable to afford wargames on my meagre pocket money - I began building wargames for myself.

In many ways, this was a rather quixotic enterprise. My rules made the games virtually unplayable; the maps were badly prepared and tatty; and the historical background was not terribly accurate. But the games - crude simulations of the battles of Arnhem and Waterloo - maintained my interest through several years, until I had the opportunity to buy 'proper' games.

However, it was not really until I got to Cambridge that I had the opportunity to play decent games with good opposition. As a result, I wasn't here long before I discovered that self-taught generals are not the best ....

Board wargames have been designed to cover almost every military action that has ever taken place. Few games constrain you to such an extent that it is impossible to design actions which never actually occurred in history. Still other games are designed to cover precisely these situations - a German invasion of Britain, in June 1940, for example. And yet others are wholly speculative - there are many games set in the far, star-faring future, including 'Star Fleet Battles', which simulates combat between Star Trek-like space-ships. The possibilities are limitless.

Units in wargames are, inevitably, as varied as the subject matter of the game. In a small-scaled game, such as 'Squad Leader', individual soldiers are represented by their own counters. Games related to military actions cover units of all sizes up from this - through platoons, battalions, divisions, up to the rather abstract armies of 'Diplomacy'. Again, other subject matters call for other units. Games of naval combat will have counters representing aircraft carriers and battleships; games of aerial combat have squadron counters; games set in space might have Death Stars, or colony transports.

Each unit will be described by a set of factors. These will usually be printed on the cardboard counters representing the units, or occasionally will be detailed in the rules. In the simplest cases, these are likely to be just a combat factor and a movement factor. The combat factor determines the effectiveness of the unit in attacking or defending against other units. The movement factor determines the rate which the unit can move on the board - tanks, for example, move faster than infantry. In more complex games, further factors are added - the range of the unit's weaponry, separate attack and defence factors, type of armour, unit designation, and so forth.

The board itself will frequently be marked in hexagons (hexes). This is not for some obscure statistical or trigonometric reason: simply that the distance from one hex to the six adjacent ones is the same in all six directions. On a squared board, those squares 'diagonally adjacent' could either be one or two steps away.

Movement on the map-board is controlled by a variety of terrains represented by different markings and colours. The nature of terrain represented again varies with the subject of the game. A simulation of a desert war would not have much call for symbols for marshland on the map-board; a naval map-board might not have any terrain other than blue sea.

Different units will usually be affected in different ways by this variety of terrains. Tanks, for example, will find their movement hampered by forest to a greater extent than foot-soldiers or aircraft. The effects of the terrain are often summarized in a chart known (imaginatively) as the Terrain Effects Chart (TEC).

Another important chart provided with most games is the Combat Results Table. When opposing units meet, they will frequently engage in combat, and the CRT is the method for determining the outcome. The ratio of the opposing forces' strengths is determined, and a dice thrown, to represent the element of chance present in all conflicts. These values are cross-referenced on the CRT, and the result of the combat is determined. The nature of the result is, again, highly dependent on the game. In simulations of ancient and medieval combat, the results will be fairly blood-thirsty - cavalry hit by cannon fire is pretty well out of circulation. However, in larger-scale and more modern simulations, combat will more often involve retreats and step reductions in strength, rather than total destruction. Only when a unit is totally surrounded is it likely to be totally overcome.

The aim of the wargamer is usually to fulfil a set of Victory Conditions. These conditions describe strategic or tactical targets for each player - perhaps to inflict a given level of casualties on the opponent; perhaps to capture a strategically important location; perhaps to hold onto a river crossing. Close knowledge of the victory conditions is very important; wiping out an opposing army is of no use to you if the last battered battalion is left holding out at their HQ, preventing you from fulfilling your victory conditions.

It is somewhat inevitable that, as the games become more sophisticated and designed to cover all eventualities, the rulebook will become proportionately heavier. In this way, wargames compare unfavourably to Chess and Go: however, many of the most complex games have programmed instructions, enabling you to start on basic scenarios before you have worked through all of the rules.

An example of this, 'Blitzkrieg', manufactured by Avalon Hill, has a set of Basic Rules, which enable you to start playing after maybe an hour's reading and setting up. As you become more familiar with the game, more rules are added - supply; stepped reduction in unit strength; air operations; industrial resources; amphibious landings; neutral countries producing militia units. By the time you are playing all of the rules, the game is probably as close to real life as it can be without breaching the Official Secrets Act ....

Interest in board wargames may come from several directions. For some people, there is an interest in the relative performance of units in battle. How would a British infantry platoon fare against a German tank unit in the Battle of Normandy, after D-Day? Wargaming enables such questions to be answered in a way which is both interesting, and as realistic as the player chooses to make it.

The re-running of battles and the staging of hypothetical conflicts is another area of interest to some. For example, the Japanese were said to have made tactical errors in the Battle of Midway, which cost them victory. Could the battle have gone the other way? Another example mentioned earlier is the hypothesized invasion of Britain in 1940 - Operation Sealion. Would a German victory in the Battle of Britain have tipped the balance in their favour?

A third reason for wargaming is simply for the intellectual challenge - the same reason why many people play Chess. It could be argued that luck will have a greater significance than in Chess or other games of skill: but in most balanced games, it will be the quality of the generalship which will decide the game.

Finally, but certainly not least in importance, wargaming can be a social activity - especially if you are not *too* concerned about winning ! There are a large number of multi-player games - from 'Dune' to 'Diplomacy'; from 'Cold War' to 'Machiavelli'. In these games, the interaction between the players becomes significant, as well.

Games such as these are particularly favoured by CUBWS - the University Board Wargaming Society. The thinking behind this is that, if you want to play a two-player game, you can always get together with somebody during the week for a game (and if you can't find an opponent, the newly-inaugurated 'Opponents Wanted' system might help you ....). But organising six people for a game of 'Machiavelli' - a Diplomacy-style game set in Renaissance Italy - is not so easy. Our meetings on Sundays from 10 am in Emmanuel Upper Hall aim to provide enough players in one place to allow the various multi-player games to be played. If this article has sparked any interest in board wargaming within you, feel free to contact either me at Christ's, Richard Brodie at Emmanuel, or Chris Doran at Jesus, and we will send you more information.

As for the supply of wargames, Cambridge is lucky enough to have an excellent little shop, 'Games & Puzzles', which is situated just behind Sainsbury's. It sells a wide variety of wargames and role-playing games at a very competitive price. Most conflicts from history are covered in at least one game, so you are almost certain to find something to your taste.

But even if you find the best the market has to offer does not satisfy you, you can always build your own game. And if no-one is willing to play your platoon scale simulation of the Iran-Iraq war '79-'83 with turns representing two hours - well, even playing solo is an enlightening experience ....

ACADEMIC FREEDOM - A PERSONAL VIEW OF LIFE BEYOND THE PhD.

John Skilling (St. John's College, Cambridge)

I suffer from a delusion or a dream, an illusion or an ideal. I happen to believe in the life of the mind, in free intellectual enquiry, and in the open dissemination of new knowledge and new ideas. I am an optimist.

Looking at our brilliant and complex civilisation, I see its guiding spirit and primary cause as intellectual freedom. It is free enquiry and free expression which have lifted us away from primitive poverty and given us powers which are literally beyond the dreams of our forebears. I think it no accident that the historic breakthrough from superstitious ignorance occurred in the (relatively) liberal West, and I count myself fortunate indeed to have been born in such a place at such a time. Intellectual freedoms are to be found in a myriad places in our society, but it is our Universities which lay long and formal claim to them, through their statutory aims of "education, religion, learning and research". Here, in Cambridge, our Colleges also lay such claim.

Twenty years ago, when I was starting my academic career, I understood all this as a sort of litany, as a familiar but perhaps rather empty church service. My research was interesting enough in its own way, adequate in quality to allow my appointment as a Lecturer, and to publish the usual number of papers, but not particularly important. It was what I have come to define as "orthodox" research. Rightly, most research, most of the time, is orthodox.

Now one of the great advantages of a Collegiate system is that it allows and encourages random conversations with colleagues in other disciplines. In the nature of things, this is usually just ordinary chit-chat, discussing political issues, or whether a particular student is working hard enough (!), or whether Dr. X's wife has had her baby yet. But one lunch in St. John's in September 1977, I discovered that a colleague had stumbled across an idea that was actually important.

"A new method of interpreting incomplete noisy data in radio interferometry" sounds ferociously dull, and indeed rather more than 99% of papers with such titles are ferociously dull. Yet my colleague had realised, as I too came to see, that this new idea was very general and powerful. We call it "maximum entropy", which sounds deep, and is. Its applications have led us into a host of fields, from chemical spectroscopy to medical imaging, from the insides of nuclear reactors to the insides of proteins, from astronomy to geophysics, and each of us round the world and across the Atlantic more often than either of us can remember. Its practice has led us deep into computer architectures and algorithms. Its theory has led us into the most fundamental areas of inductive reasoning and statistical thought. Indeed, we are literally concerned with how we are *supposed* to infer knowledge of the complex world around us. Far from being "orthodox", this research has become "passionate".

I owe my Department considerable gratitude. It allowed me to drop my previous work and do what I considered to be important. This is what academic tenure is about. Tenure allows one to do what one wants. Outside Universities, few understand the importance of this, seeing the privilege but not the responsibility, and the institution seems doomed. Tenure gives freedom. Of course, freedom can be abused, though very few do abuse tenure. More often, tenure is used properly, to let one do what is seen to be important. One may not gain promotion, but at least one keeps one's basic job.

What of teaching, though? After all, education is the first statutory function of the University. It also happens to be fun: I like lecturing and supervising. Mathematics is full of extraordinarily intricate and beautiful patterns of logic. I enjoy trying to show these to students - not just spraying symbols on a blackboard to be erased four minutes later, but trying to show something of the subtle rationale behind them, and to nurture the natural enthusiasm of young, fresh minds.

Well, of course teaching takes time. A sixty-hour working year may not sound much for a lecturer, but that bald statistic obscures the truth. It ignores preparation of the course, example sheets, liaison with the other lecturer, discussion of the schedules, correspondence, committee meetings to discuss this or that, perhaps a Board to attend periodically, miscellaneous student enquiries, odd little responsibilities,... Examining is another part of the job, involving setting, organising, checking, discussing, corresponding, marking, publicising, reporting,... Relative to the productive "contact hours" of teaching, a phenomenal amount of effort is dissipated in organisation. For example, committees and departments can easily spend more than 24 man-hours discussing schedule changes for a 24-lecture course, which begins to feel unbalanced. College teaching also takes time, with its standard termtime load of six hours a week more than doubled by preparation and organising if one takes the job seriously.

The end result is that a lecturer's working week is shattered, and there is never any period of coherent time in which to do fundamental thinking. With orthodox research, that does not matter. Orthodox research can be put off for months on end, if need be. Passionate research can not. I would not, I think, have chosen it thus, but at that level of commitment I have found that the call of the mind, like that of the running tide, "is a wild call and a clear call that may not be denied".

In my own case, the conflict came to a head last winter, in the early months of 1987. I had almost cleared my desk of its endless backlog of necessary tasks, and would have done so but for the necessity of setting a variety of Part II questions, when the Lent term hit me. I will not easily forget the catalogue of disasters and near misses which befell during that term.

It became distressingly clear that proper professional attention to University and College is flatly incompatible with passionate research. It should not be so, but it is. Something had to give. Family? - of course not. Research thinking? - probably impossible psychologically. College? - the traditional sacrifice for over-worked lecturers. Yet nearly all the multi-disciplinary breadth of my research happens to have been mediated through my College: unusually, much less has come through my University department. University? - the institution which taught me so much, granted me tenure, and each month paid my major salary. Yet it was the extra little University duties, each fairly small in itself, which were destroying the coherence of my time, and draining the very lifeblood of thought which the University claims to encourage. Accordingly, I prepared to resign the University office I had held with quiet pride for fifteen years. It was a difficult decision, even though consultancy work could be expected to take care of the financial cost.

A more experienced colleague suggested that instead of straight resignation, I might seek a period of unpaid leave, and with nothing to lose, I did so. To my pleasure and slight surprise, both my Faculty Board and the General Board approved the proposal, so that I am now starting three years without my salary, but with the freedom to do what I know in my heart I ought to do. Being fortunate enough to retain my Fellowship, I can also keep doing some of the teaching I enjoy.

Now that the first term of freedom is over, I know I did right. A new idea arrives unexpectedly, and I have time to concentrate on it. This is what a University is *for*. I'm not being lazy - in fact I am working at least as productively as I have ever done, and the sense of relief at doing things *well* is enormous. A friend commented the other day that I seemed to be a full two inches taller since last winter, and I know what he means.

I can also find a little time to write an essay like this one, hoping to illuminate for students something of the quality, intensity, and difficulty of life beyond the PhD. So, students of the next generation, don't expect life to be easy. It isn't, and as academics we live in paradoxically difficult times. Do aim for life to be rewarding. It can be. And lastly, keep faith with your own skills. Even for a mathematician, it is simply not true that one is past it at the age of 30!

DIFFERENTIATION RULES, OKAY

Clive Monk (Queens' College, Cambridge)

The purpose of this article is to take a light-hearted look at how rules for differentiating a function can be thought of as "saying" something else, totally unconnected with differentiation. It is the chain rule and rule for differentiating a product (once) which we will consider. Detailed proofs of the results claimed are not given although it is hoped that enough information has been supplied to enable the interested reader to construct these for him/herself.

Hopefully the notation introduced is fairly self-explanatory. We mention the following now to avoid interruption later :

$$!n = \prod_{m=1}^{n} m, \text{ often denoted } \underline{n}! \text{ or (sadly for historical reasons) } \Gamma(n+1)$$

This function can be extended in a natural way to (most) complex numbers z (although we will not use this fact). All the well-known formulae for $\Gamma(z)$ turn out to be even simpler when considered as formulae for !z (to spell it out, people have been studying the "wrong" function!) For an interesting list of other historical "mistakes", see Eagle[1].

$$p^m \rfloor \sum_{i=0}^{\infty} a_i p^i = a_m$$

ω = the smallest infinite ordinal

The idea is that we have a "function" of x which we can differentiate indefinitely using the above rules. We differentiate it a large number of times and then, as a huge anticlimax, we put x=0. This is all that we are required to be able to do, so our "function" needn't be a function at all. For example it is an easy consequence of the techniques discussed that if

$$\left[\frac{d^n}{dx^n} C\right]_{x=0}$$

gives the number of connected graphs on n (≥0) labelled vertices then

$$e^{C-1} = \sum_{n=0}^{\infty} \frac{2^{\binom{n}{2}} x^n}{!n}$$

which must be thought of as a "formal object" as it has zero radius of convergence! A natural first candidate is

$$e^{e^{x}}$$

For $x \in \mathbf{C}$ this is an analytic function, although it is interesting to consider what e^x "is" for other x. If x is d/dy, for example, then e^x is best described as "the act of replacing y with y+1". At this point the reader is likely to be interested in raising e to the power of other things. Might I suggest a skew-symmetric matrix?

Anyway, we differentiate this function (whatever this might "be") n times with respect to x and then seemingly destroy all our hard work by putting x=0. What do we get? ke, where $k \in \mathbf{Z}^+$ (to offend no-one : some people have strong views as to whether 0 should be a natural number). So, what *is* this k which for large n we seem to have to work so hard to obtain? Well, k is the number of ways of grouping n distinct objects as made clear below for the case n=3, where k=5.

(1) (2) (3)
(1)(3) (2)
(1) (2)(3)
(1)(2) (3)
(1)(2)(3)

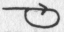
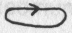
To see this, take the objects to be the differentiations. After differentiating m times, terms take the form

$$e^{e^x} e^x \dots e^x$$

where there are at most m e^x s. Each e^x corresponds to a group of objects (differentiations). Differentiating such an e^x corresponds to placing the m+1th differentiation with this group. Alternatively the very first term may be differentiated corresponding to starting off a new group. Finally x is put equal to zero which has the effect of enumerating the possibilities. Thus

$$\left[\frac{d^n}{dx^n} e^{e^x}\right]_0 \text{ gives } k(n)e \text{, i.e.}$$

$$k(n) = \left[\frac{d^n}{dx^n} e^{e^x - 1}\right]_0 \qquad (1)$$

Number theorists will note that if we are only interested in k(n) mod m, we may replace m with 0 while differentiating. Thus we pretend that $e^{mx} = 1$. Since there are then only finitely many possibilities, k(n) mod m must do this :  . A moment's more thought (e.g. using matrices) reveals that in fact it does this :  . For example, $k(n+13) \equiv k(n) \pmod 3$.

The technique of "pretending" that something is something else is common in mathematics. For example Goodstein's proof[2] of a theorem about the natural numbers works by replacing certain natural numbers in an expression by ω. Also there is the process of Godel numbering[3] and that of pretending an algorithm is a number - useful for proving the insolubility of the halting problem[4]. Hardy and Wright[5] give a more sophisticated example of how it can be useful to identify functions which would normally be considered distinct ($(1-x)^{-5}$ and $(1-x^5)^{-1}$) as well as showing "rewards" (e.g. theorems 362 and 363) for pretending a partition is a power of x. (1) suggests the simpler

$$\left[\frac{d^n}{dx^n}\,(e^x-1)^m\right]_0$$

This gives the number of ways of putting n distinct objects into m distinct (=labelled) containers such that no container is empty, and is often denoted $S_n^{(m)}$ (the Stirling numbers of the second kind^6). As before take the objects to be the differentiations. The containers may be layed neatly in a row, thus : $(e^x-1)\ldots(e^x-1)$, and we abuse the English language by "confusing" "differentiating a container" with "placing a differentiation inside a container". Each container must be differentiated at least once to avoid a zero appearing in the product.

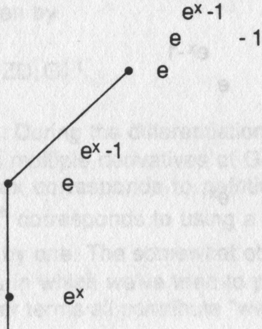
Having turned our attention from e^x to the less natural e^x-1 we are in a better position as regards iterates ($0\to0\to0\to\ldots$ is easier to cope with than $0\to1\to e\to\ldots$). Let $f_i(x)$ be the ith iterate. We know about f_1^m and f_2 (which gives k(n) except when n=0 : we will take $n\geq1$ from now on). We consider next f_3 and leave the reader the (easy) task of generalizing to f_i^m etc.

$$\left[\frac{d^n}{dx^n}\,f_3\right]_0$$

gives the number of rooted trees with n labelled endvertices (leaves) in which each leaf is a distance of 3 from the root. To see this we write the terms in a rather unconventional way (as a tree) while we perform the differentiations. This is best made clear by an example. After the first differentiation we get

$$e^{e^{e^x-1}-1}\qquad e^{e^x-1}\qquad e^x$$

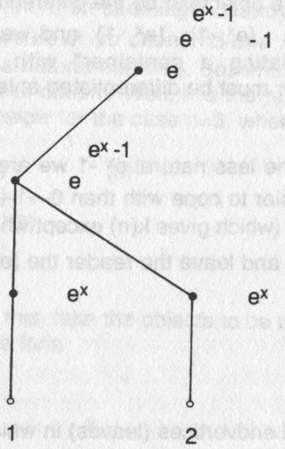
which we write as



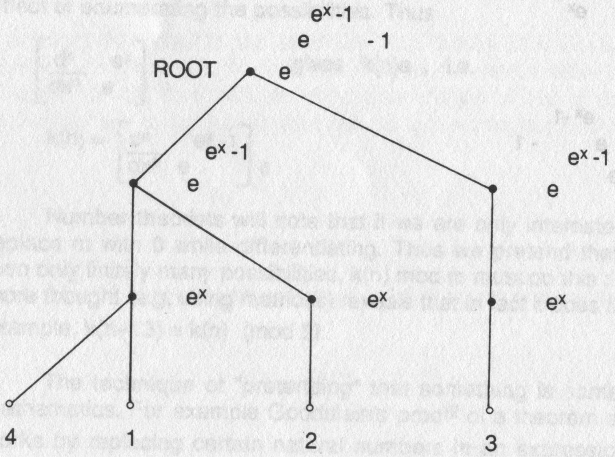
and understand that the term is the product of the "values" of the dark nodes. For our second differentiation there are three nodes we could choose to differentiate (= make a branch grow out of). Suppose we choose the

e^{e^x-1} node. $\frac{d}{dx}e^{e^x-1} = e^{e^x-1}e^x$

so we get


$e^{e^{e^x-1}-1}$
e^{e^x-1}
e^x
e^x
1
2

By now the reader has probably got the general idea. We make two more differentiations to illustrate the process further.


$e^{e^{e^x-1}-1}$
ROOT
e^{e^x-1}
e^{e^x-1}
e^x
e^x
e^x
4
1
2
3

which is a term in

$$\frac{d^4}{dx^4}e^{e^{e^x-1}-1}-1$$

Note that, when x=0,

$$e^x = e^{e^x - 1} = e^{e^{e^x-1}-1} = 1$$

so all terms are given "weight" 1. The upshot of all this is that

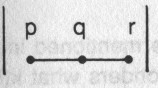
$$\left[\frac{d^n}{dx^n} f_3\right]_0$$

is as claimed. As an example of what can be done by inserting a second variable,

$$\left[\frac{\partial^m}{\partial y^m}\frac{\partial^n}{\partial x^n}\left(e^{e^{y(e^x-1)}-1}-1\right)\right]_{x=y=0}$$

gives the number of rooted trees again of "height" 3 and with n labelled leaves, but this time there are precisely m labelled "twigs" (vertices which support the leaves) as well (m, n not both 0).

We take a break from counting trees to consider chromatic polynomials, an introduction to which can be found in, amongst other places, Eureka[7], Wilson[8], and many other books on graph theory. Take any graph G (with no loops or multiple edges) and identify each vertex with a different symbol. Let |G| denote the sum of all possible products of mutually unconnected vertices. For example



$$= 1 + p + q + r + pr$$

Let D denote the operation of differentiating w.r.t each vertex in turn, and Z that of putting them all equal to 0. Then, for t a new symbol, the chromatic polynomial of the graph is given by

$$G(t) = ZD|G|^t \qquad (2)$$

Explanation: During the differentiation process each term is the product of a $|G|^{\text{colours left}}$ and various multiple derivatives of G which we will call colours. Differentiating a colour w.r.t a vertex corresponds to painting the vertex with this colour. Differentiating the $|G|^{\text{colours left}}$ corresponds to using a new colour and causes the number of colours left to go down by one. The somewhat obscure definition of |G| insures that, on applying Z, those terms in which we've tried to paint two adjacent vertices the same colour vanish and the other terms all contribute "weight" 1.

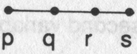
Improvements are possible for "special" G. For example if G is $K_{m,n}$ (see Wilson[8], or work it out from the context) then the chromatic polynomial is given by

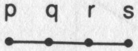
$$G(t) = \left[\frac{\partial^m}{\partial y^m}\frac{\partial^n}{\partial x^n}(e^x + e^y - 1)^t\right]_0 \qquad (3)$$

for similar reasoning. We mention in passing that for fixed t, $ZD(|G| - 1)^t$ gives the number of colourings in which all t colours are used. We next look at

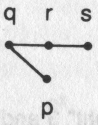
$$ZDe^{|G|-1} \qquad (4)$$

(note the -1 as with the trees). This gives the number of ways of "folding G up". The proof mimics that of (1) together with the observation that the nature of $|G|$ -1 ensures that only groupings in which adjacent vertices have not been identified get counted. We denote the number of ways of "folding G up" by k(G) and note that if G has no edges then k(G) = k(n), where n is the number of vertices of G (perhaps 0). For example,

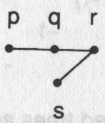
k() = 5 , namely



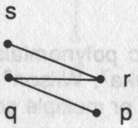
,



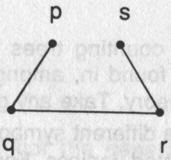
,



,



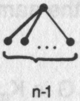
,



Also $k(K_n) = 1$ and $k(K_{m,n}) = k(m)\,k(n)$ $(m,n \geq 0)$. The graphs K_n are mentioned in an interesting article by Graham Nelson [this issue of Eureka]. One wonders what $k(P_n)$ and $k(C_n)$ (where P_n and C_n are as in the above article) are. Does their being "special" allow us to improve (4) in a similar way to which we improved (2) to get (3)? Note that we've already seen that $k(P_4) = 5$. To find out, note that

$$k(G) = \frac{1}{e} ZD \sum_{t=0}^{\infty} \frac{|G|^t}{!t} = \frac{1}{e} \sum_{t=0}^{\infty} \frac{G(t)}{!t} \qquad (5)$$

expressing k(G) in terms of G(t) alone. Now the chromatic polynomial for a tree with n vertices is $t(t-1)^{n-1}$ which does not depend on the tree's structure. Hence if G is *any* tree on n≥1 (labelled) vertices (in particular P_n), then

$k(G) = k($  $) = k(n-1) = \left[\frac{d^n}{dx^n} P\right]_0$

where

$$P = \int e^{e^x - 1}\, dx$$

Suppose

$$k(C_n) = \left[\frac{d^n}{dx^n} C \right]_0$$

An exercise in Wilson[8] tells us that $C_n(t) = (t-1)^n + -^n (t-1)$ from which, using (5),

$$C = \sum_{n=0}^{\infty} \frac{x^n}{!n} \frac{1}{e} \sum_{t=0}^{\infty} \frac{(t-1)^n + -^n (t-1)}{!t}$$

$$= e^{e^x - x - 1}$$

Hence $k(C_n)$ is the number of ways of grouping n distinct objects such that each group contains at least two objects. We can modify (5) to obtain a neat formula for k(G) in terms of the coefficients of G(t). Suppose

$$G(t) = \sum_{n=0}^{d} a_n t^n$$

Then

$$k(G) = \frac{1}{e} \sum_{t=0}^{\infty} \frac{G(t)}{!t}$$

$$= \left[\frac{1}{e} \sum_{t=0}^{\infty} \frac{\left(e^{d/dx}\right)^t}{!t} G(x) \right]_0$$

$$= \left[e^{e^{d/dx} - 1} G(x) \right]_0$$

$$= \left[\sum_{n=0}^{\infty} \frac{k(n)}{!n} \left(\frac{d}{dx}\right)^n G(x) \right]_0$$

$$= \sum_{n=0}^{d} k(n)\, a_n \qquad (6)$$

For example icosahedron(t) = $t^{12} - 30t^{11} + 415t^{10} - 3500t^9 + 20023t^8 - 81622t^7 + 241605t^6 - 517360t^5 + 780286t^4 - 782108t^3 + 463310t^2 - 121020t$,

so there are

$1 \times -121020 + 2 \times 463310 + 5 \times -782108 + 15 \times 780286 + 52 \times -517360 +$
$203 \times 241605 + 877 \times -81622 + 4140 \times 20023 + 21147 \times -3500 + 115975 \times 415 +$
$678570 \times -30 + 4213597 \times 1 = 26793$ ways of "folding up" the icosahedron graph.

Now

$$k(n) = \sum_{m=0}^{\infty} \frac{S_n(m)}{!m} \left(= \sum_{m=0}^{n} \frac{S_n(m)}{!m} \right)$$

In fact

$$\sum_{n=0}^{d} \frac{S_n(m)}{!m} a_n$$

has a nice combinatorial interpretation as the number of ways of folding G to a graph with m vertices. For the latter is

$$p^m \text{⌡} ZD\, e^{P(|G| -1)} \qquad \text{(P a new symbol)}$$

and a similar calculation to that for (6) shows that this equals the former. So far we have relied on spotting functions which behave interestingly when differentiated. The next example is more subtle and shows how it is possible to start out with a combinatorial question and solve differential equations to get the function you want.

As motivation for the problem tackled consider $y = (1 - x)^{-1}$ (because, like e^x, there is a nice formula for its nth derivative). $dy/dx = y^2$ so a natural activity to do while differentiating is to draw labelled trees with a y at each vertex. Since $y(0) = 1$, each tree drawn will have "weight" 1. Differentiating a vertex corresponds to shining sunlight on that part of the tree. Not all labelled trees will be produced (the correct function for doing this satisfies

$$\frac{dy}{dx} = e^{x\, dy/dx}$$

an easy consequence of Stanley[9]. This superb paper contains many other interesting examples. A by no means apparent recurrence relation for the number of n-square polyominos in which each "row" is an unbroken line of squares is obtained. Also, anyone interested in a combinatorial interpretation of

$$\left[\frac{d^n}{dx^n} (1-x)^{-x} \right]_0$$

should look at example 6.6. The author points out that "this example was contrived solely to obtain the correct answer" (of $(1-x)^{-x}$). However the interpretation, too, is surprisingly elegant. We call a tree which is produced an *alert* (for Ascendingly LabellEd Rooted Tree) because it will be useful to call it something and I can't think of anything better to call it. Like the real thing, alerts "grow" in some kind of order; leaves do not form until there are twigs ready to support them!

We now pretend it was alerts which we were interested in all along and ask how many alerts have n vertices and m leaves. At each vertex which isn't a leaf we write y and at the leaves we write z. We demand that

$$y(0) = 1 \qquad z(0) = p \tag{7}$$

and take the "value" of a tree to be the product of the vertex values when x=0. Thus, if we can find a suitable f to generate the alerts, then the number required will be

$$p^{m}) \left[\frac{d^{n}}{dx^{n}} f \right]_{0}$$

We allow our alerts to grow :

$$\frac{dy}{dx} = \frac{dz}{dx} = yz \tag{8}$$

(Note that, unlike the real thing, leaves can turn into branches!). The root is not considered a leaf (for if it was, would have to get value z^2). Solving (7) and (8) gives

$$y = \frac{(1-p)e^{px}}{e^{px} - pe^{x}} \qquad z = \frac{p(1-p)e^{x}}{e^{px} - pe^{x}} \tag{9}$$

so that $f = 1 + x - \ln(e^{px} - pe^{x})$. From (9) it can be seen that the number of alerts with n vertices and m leaves is

$$\sum_{i=0}^{m} -^{i} \binom{n}{i} (m-i)^{n-1} \qquad (m \geq 0, n \geq 1, 0^0 = 1)$$

We've considered only trees which have been labelled in some way. If T_n is the number of unlabelled trees on n vertices with one "distinguished" vertex (root) then

$$T(x) = \sum_{n=1}^{\infty} T_n x^n$$

satisfies the remarkable relation

$$T(x) = x e^{\sum_{n=1}^{\infty} n^{-1} T(x^n)}$$

This follows from

$$\sum_{n=1}^{\infty} T_n x^n = x \prod_{n=1}^{\infty} (1-x^n)^{-T_n}$$

a proof of which may be found in Harary[10] (hint : $(1 - x^n)^{-T_n} = (1 + x^n + x^{2n} + \dots)^{T_n}$ and compare coefficients). Interested readers should also see the Encyclopedia Brittanica[11]. Finally I mention the "opposition". Many interesting combinatorial facts can be proved[12] *without using a single generating function!* Some may hold the view that problems are hard enough as it is without attempting them "with one's hands tied behind one's back". However I feel that this is a useful thing to do as it helps one realise when generating functions are being used merely through habit.

REFERENCES

1 Eagle, A. The elliptic functions as they should be, Galloway and Porter
2 Henle, J.M. An outline of set theory, Springer-Verlag
3 Hofstadter, D.R. Godel, Escher, Bach : An Eternal Golden Braid
4 Boolos, G.S., and Jeffrey, R.C. Computability and logic, Cambridge University Press
5 Hardy, G.H., and Wright, E.M. An introduction to the theory of numbers, page 288, Oxford University Press.
6 Abramowitz, M., and Stegun, I.A. Handbook of mathematical functions, page 824, Dover
7 Eureka 44, The Archimedeans
8 Wilson, R.J., Introduction to graph theory, Longman
9 Stanley, R.P., "Generating functions", from Studies in Combinatorics (vol. 17), page 136, Math. Ass. America
10 Harary, F., Graph theory, Addison-Wesley
11 Encyclopedia Brittanica (under Combinatorics)
12 The Mathematical Intelligencer, vol. 5, no. 4, page 21, Springer-Verlag

RAMANUJAN - A GLIMPSE OF HIS LIFE AND HIS MATHEMATICS

Béla Bollobás (Fellow of Trinity College, Cambridge; Reader in Pure Mathematics)

(This article is based on talks given to the Trinity Mathematical Society on 9th June, 1987, and to the Adams Society on 19th January, 1988).

The history of mathematics is rich in exciting and peculiar geniuses. But none was more extraordinary than Srinivasa Ramanujan, the great Indian mathematician. In vague terms his romantic but tragic life is quite well known, but the details are shrouded in mystery even now, and most mathematicians would be hard pressed to state even one of his theorems. In this brief note we shall sketch the story of his life, including some hitherto less known facts, and give a short introduction to the kind of mathematics he did.

In the 19th century mathematical analysis in England was in a sorry state: on the Continent, especially in France and Germany, revolutionary advances were made, while England failed to produce any analyst of the first rank. However, by the 1930s England could boast of a flourishing school of analysis, mostly due to the efforts of two outstanding mathematicians, G.H. Hardy and J.E. Littlewood.

The Hardy-Littlewood partnership, perhaps the greatest scientific partnership ever, dominated English mathematics for several decades. Hardy, born in 1877, was the older, by 8 years. Their collaboration began in 1911 and their last joint paper was published in 1948, a year after Hardy's death. For most of this period they were fellows of Trinity College, Cambridge.

In 1913 Hardy received an unsolicited letter from India.

> Madras, 16th January 1913
>
> Dear Sir,
>
> I beg to introduce myself to you as a clerk in the Accounts Department of the Port Trust Office at Madras on a salary of only £20 per annum. I am now about 23 years of age. I have had no university education but I have undergone the ordinary school course. After leaving school I have been employing the spare time at my disposal to work at Mathematics. I have not trodden through the conventional regular course which is followed in a university course, but I am striking out a new path for myself. I have made a special investigation of divergent series in general and the results I get are termed by the local mathematicians as 'startling'.
>
> ... I would request you to go through the enclosed papers. Being poor, if you are convinced that there is anything of value I would like to have my theorems published. I have not given the actual investigations nor the expressions that I get but I have indicated to the lines on which I proceed. Being inexperienced I would very highly value any advice you give me. Requesting to be

excused for the trouble I give you.

I remain Dear sir Yours truly

S. Ramanujan

P.S. My address is S. Ramanujan, Clerk Accounts Department, Port Trust, Madras, India.

The 'enclosed papers' contained a long list of results claimed by the writer. Apparently this letter was read by Hardy and Littlewood after hall. Some of the results were well known and some they could prove, although not too easily, some were clearly incorrect and some could be proved to be incorrect, but several looked correct and still baffled them entirely. The entire approach was permeated by a certain naïveté, as remarked by Littlewood. "I have a vague theory as to how his mistakes have come about. I imagine that he is satisfied if he can convince himself that his results are correct & he has probably staked on certain operations on dgt. [divergent] series involving primes being legitimate. His results are just about what one wd [would] get if $\zeta(s)$ had no zeros in $\sigma>0$."

Although the basic mistakes committed by this unknown Indian seemed very peculiar to Hardy and Littlewood, the beauty and strength of many of the other results claimed were indeed startling. To quote Littlewood again, "... this appears to be correct, though I haven't verified all the algebra. Not vy [very] exciting of course, though I shd [should] say rather a tour de force without Cauchy's theorem." "17 is an exp [expression] for $\vartheta(0, \sqrt{210})$ isn't it? Surely this & 20-23 are new & if so, vy [very] exciting? I can believe that he's at least a Jacobi. How maddening his letter is in the circumstances. I rather suspect he's afraid that you'll steal his work."

As this passage shows, Hardy and Littlewood came to the conclusion that the writer of the letter was not a lunatic, nor was he just an amateur, but a man of most unusual talent. To be compared to Jacobi, the great German master of formulae was tremendous praise indeed. And to have arrived at that level without any professional training was quite unbelievable.

So who was the author of this letter? Srinivasa Ramanujan was born into a very poor Brahmin family in 1887, probably on the 22nd December, in Erode, south of Madras. [This is his accepted date of birth, although it is a little peculiar that he himself wrote in January 1913 that he was about 23, and later, in the Admissions Book at Trinity College, he gave his date of birth as December 1888.] His father was an accountant and his mother was an intelligent woman, who was literate in Tamil and also knew some Hindu astronomy. As he was extremely poor, Ramanujan was educated as a free student in the High School at Kumbakonam.

While in this school, he came across a rather peculiar book, entitled *A Synopsis of Elementary Results in Pure Mathematics*, published in two volumes in 1880 and 1886. This was the work of a certain George Shoobridge Carr, a private coach in London and formerly Scholar of Gonville and Caius College, Cambridge, and contained some 6000 assertions with 'abridged demonstrations'! Carr was a rather peculiar man himself, who came to Cambridge to read mathematics when he was nearly forty, and in 1880, the year he published the first volume of his *Synopsis*, was 12th Senior Optime [12th in the Second Class]. Nevertheless, the book showed signs of real scholarship and a glowing love of mathematics.

Today, the only claim to fame of Carr's *Synopsis* is that it left a profound impression on Ramanujan. Ramanujan set himself the task of establishing all the formulae in the book and after this mammoth enterprise he never ceased to excell in finding beautiful and unexpected formulae. This extraordinary feel for formulae and formal manipulations was to be his main strength.

Ramanujan passed the Matriculation Examination of the University of Madras in 1903, and was awarded a scholarship at Kumbakonam. Unfortunately his smooth academic progress did not last long. In 1904 he became temporarily insane, left home and for months on end wandered about, getting 700 miles from home. Later he couldn't remember how he had lived. In 1907 he sat for the First Examination in Arts but failed in all subjects, including mathematics. Two years later he had a very serious operation that kept him in bed for months. But he continued working on mathematics.

In 1909 Ramanujan married Srimathi Janaki, a 9 year old girl. In order to support himself and his wife, he became a clerk in the Port Trust Office in Madras. In his spare time he started genuine research in mathematics, working on elliptic functions and elliptic integrals, continued fractions and the distribution of primes.

Hardy was not the first man in England approached by Ramanujan. First he wrote to H.F. Baker and then to E.W. Hobson, but they ignored his letters. In later years Littlewood often chuckled over the embarrassment of his colleagues who failed to recognise a genius. But Hardy did, and a long, although occasionally strained correspondence followed. First the University of Madras granted Ramanujan a special scholarship of £60 per annum for two years at the instance of Sir Gilbert Walker, Head of the Meteorological Department in India, formerly Fellow and Lecturer in Mathematics at Trinity College. This sum was sufficient to enable a married Indian to live in tolerable comfort. Later Hardy, and another Trinity mathematician, E.H. Neville, managed to bring Ramanujan to Cambridge. Ramanujan got a very generous scholarship of £250 from Madras, of which £50 was allotted to support his family in India, and an exhibition of £60 from Trinity.

Leaving India was a momentous step for Ramanujan: as a Brahmin, he was not supposed to do so. However, eventually Ramanujan obtained the permission of his mother, got the consent of Ramagiri, his favourite Indian goddess, cut his hair, and on March 17, 1914, in the company of Neville, sailed for England, leaving his young wife behind.

Ramanujan was not very well prepared for Cambridge and arrived at a rather unfortunate time. In many ways, he had less mathematical sophistication than most undergraduates today. As Littlewood wrote later, "the clear-cut idea of what is meant by a proof, nowadays so familiar as to be taken for granted, he perhaps did not possess at all; if a significant piece of reasoning occured somewhere, and the total mixture of evidence and intuition gave him certainty, he looked no further." He hardly knew Cauchy's Theorem, and practically never used it. Although Ramanujan considered his work on the classical theory of numbers very important, he knew very little of the general theory of arithmetical forms and even less about the theory of analytic functions, in particular, about the zeta function of Riemann, upon whose properties the theory of primes depends. However, in formal manipulations Ramanujan was truly exceptional. To quote Hardy, "It was his insight into algebraical formulae, transformation of infinite series, and so forth, that was most amazing. On this side most certainly I have never met his equal, and I can compare him only with Euler or Jacobi. He worked, far more than the majority of modern mathematicians, by induction from numerical examples; all his congruence properties of partitions, for example, were discovered in this way."

Ramanujan's teaching in Cambridge was to be shared by Hardy and Littlewood, but the War soon broke out and Littlewood left Cambridge. Ramanujan worked with great enthusiasm, having found a friend and teacher in Hardy, on some days taking half a dozen new theorems to him. As Hardy wrote, "His mind had hardened to some extent, and he never became at all an 'orthodox' mathematician, but he could still learn to do new things, and do them extremely well. It was impossible to teach him but he gradually absorbed some new points of view."

In Cambridge Ramanujan led a very irregular life. As a rather strict Brahmin, he was not only a vegetarian but had to do his cooking himself. He worked feverishly, occasionally putting in 30 hours at a stretch and then sleeping for 20. He never ceased to be exceptionally productive.

Unfortunately, in the spring of 1917 Ramanujan began to feel unwell. That summer he went to the Nursing Home in Cambridge, and from then on he was hardly ever out of various sanatoria. His illness was never clearly diagnosed: perhaps he had tuberculosis, as most commonly believed, or perhaps he had lead poisoning from the food carelessly left in his cooking utensils, as asserted by the late Professor P.A.M. Dirac. He was certainly undernourished, suffered from the cold and damp climate, and was a bit of a hypochondriac. He went from one sanatorium to another, arriving full of hope and trust, becoming disappointed soon and then trying desperately to get away. But wherever he stayed, he did mathematics with great enthusiasm.

One must sympathise with Ramanujan for there is no doubt that he suffered much. His letters to Hardy are almost exclusively about two topics: mathematics and his health. His longest stay was in the sanatorium at Matlock, where he arrived in November, 1917. Let me quote a passage from a letter of Ramanujan to Hardy written from Matlock.

> I have been here a month and I have not been allowed fire even for a single day. I have been shivering from cold many a time and have not been able to take my meals sometimes. In the beginning I was told that I could not possibly have any except the welcome fire I had for an hour or two when I entered this place. After a fortnight of stay they told me that they received a letter from you about one and promised me fire on those days in which I do some serious mathematical work. That day hasn't come yet and I am left in this dreadfully cold open room. Even if I do any serious mathematics in future I am not going to ask them for fire on that day.
>
> The bath rooms are nice and warm. I shall go to the bath room with pen and paper every day for about an hour or so and send you two or three papers very soon. This thought did not strike me before. Else I would have written something already. In a week or so you may perhaps have a complaint against me from the doctor that I am having bath every day. But I assure you beforehand that I am not going to bathe but to write something.

Although separated from her by several thousand miles, Ramanujan had trouble with his mother as well. Not surprisingly, she did not realise that her son was a genius, and that a Cambridge degree was far from being a great achievement for him. Also, it seems that she disapproved of all communications between Ramanujan and his wife. Here is a passage from another letter to Hardy from Matlock.

> It is true that I promised my mother that I was going home at the end of 2 years; I wrote them several letters $1^1/_2$ years ago that I was coming over there for the long vacation; but I had many letters of protest from my mother to the effect that I ought not to come to India till I took my M.A. degree. So I gave up the idea of going there.
>
> It is not true that I am getting letters from my wife or brothers-in-law or anybody. I had only very few formal letters from my wife just explaining to me why she had to leave my home and asking me why my mother should trouble her by [not] speaking to her when they met anywhere by chance. When I do not know the whereabouts of my wife my mother's grievance is that I have left my wife in some secret place somewhere in India and that she is waiting for me frequently to come to that place without my mother's knowledge and that I am listening to her words all the time...
>
> The initial S. in my name stands for Srinivasa which is my father's name. I haven't got a surname, really speaking.

In the autumn of 1917 Hardy put Ramanujan up for a Fellowship in Trinity College. However, he was strongly opposed by some fellows, probably mainly because of the colour of his skin, and failed to get elected. Hardy and Littlewood would not give in. They put Ramanujan up for a Fellowship of the Royal Society: a considerably higher honour than a Fellowship in Cambridge. Although this honour was richly deserved by Ramanujan, it was nevertheless a little premature since Littlewood himself had got elected only a year earlier. In February 1918 Ramanujan was duly elected an F.R.S. and so became the first Indian member of that illustrious society. As we can see from his letter to Hardy, he was overjoyed at the news.

> Dear Mr. Hardy,
>
> My words are not adequate to express my thanks to you. I did not even dream of the possibility of my election. When I opened your telegram I read thrice Fellow Philosophical Society instead of Royal Society.... Please convey my heartfelt thanks to Major MacMahon and Mr. Littlewood.

Later that year Hardy and Littlewood did manage to secure a Fellowship for Ramanujan in Trinity, but the election was still not straightforward. Many years later Littlewood wrote about the fight concerning the election. "There was much opposition. Hardy was not made an Elector, and I acted, by letter, because I was quite ill (after concussion on the top of years without proper holidays from work). I did get wind of the enemy's tactics from R.A. Herman, who was a close personal friend, and although he was against Ramanujan himself, he was always naively honest. I said: 'You can't reject an F.R.S.' 'Yes, we thought that was a dirty trick!' "

The Fellowships of the Royal Society and Trinity College, together with the financial security the latter entailed, seemed to restore Ramanujan's health. On 27 February 1919 Ramanujan sailed for home. Unfortunately the recovery did not last long. In spite of the more favourable climate and better diet, he soon fell ill again and died on 26 April.

No matter how fascinating and romantic Ramanujan's life was, his greatness lies in his mathematics. Although many of Ramanujan's mathematical achievements are easier to appreciate than the results of most other great 20th century mathematicians, in a short article we cannot hope to do more than scrape the surface. To illustrate Ramanujan's unsurpassed mastery of formulae, we shall give two examples. The first is a very simple identity, while the second is, in fact, a brief introduction to a subject, the theory of partitions, and leads on to much deeper results.

Let us see then the first example, an expression for the Euler constant

$$\gamma = \lim_{n\to\infty} \{ \sum_{k=1}^{n} k^{-1} - \log n \} = .57721...$$

Ramanujan gave the following expression for γ:

$$\gamma = \log 2 - \frac{2}{3^3-3} - 2\left(\frac{2}{6^3-6} + \frac{2}{9^3-9} + \frac{2}{12^3-12} \right)$$

$$- 3\left(\frac{2}{15^3-15} + \frac{2}{18^3-18} + \ldots + \frac{2}{39^3-39} \right) - \ldots$$

To see this, note that

$$\frac{2}{(3m)^3 - 3m} = \frac{1}{3m+1} + \frac{1}{3m} + \frac{1}{3m-1} - \frac{1}{m}$$

Consequently the right-hand side of the identity to be proved is

$$\log 2 - (1/2 + 1/3 + 1/4 - 1) - 2 (1/5 + \ldots + 1/13 - 1/2 - 1/3 - 1/4)$$

$$- 3 (1/14 + \ldots + 1/40 - 1/5 - \ldots - 1/13) - \ldots$$

so the sum of the first n terms (including log 2) is

$$
\begin{aligned}
S_n &= \log 2 + \sum_{k=1}^{n} {}^{1}/_{k} - n \sum_{k=(3^n+1)/2}^{(3^{n+1}+1)/2} {}^{1}/_{k} \\
&= \log 2 + \gamma + \log((3^n - 1)/2) + o(1) - n\{\log((3^{n+1}-1)/(3^n-1)) + O(3^{-n})\} \\
&= \gamma + n \log 3 - n \log 3 + o(1) \\
&= \gamma + o(1)
\end{aligned}
$$

Much of Ramanujan's most beautiful work was on the theory of partitions. This is a branch of mathematics that, in its elementary form, can be appreciated by everybody interested in mathematics. However, at a more advanced level it is also connected to some deep areas of mathematics, including the theory of modular forms, and has many applications in analysis, number theory, combinatorics and even physics.

A *partition* $\lambda = (\lambda_1, \ldots, \lambda_r)$ of n is a non-increasing sequence of positive integers $\lambda_1, \lambda_2, \ldots, \lambda_r$ such that $\lambda_1 + \lambda_2 + \ldots + \lambda_r = n$. The *number of partitions of* n is denoted by p(n); we call p(n) the *partition function*. Thus p(1)=1, p(2)=2, p(3)=3, p(4)=5, p(5)=7, p(6)=11 and p(7)=15. By convention one takes p(0)=1.

The *generating function* f(q) for a sequence a_0, a_1, ... is the formal power series $f(q) = \sum_{n=0}^{\infty} a_n q^n$; the function f(q) is said to enumerate the sequence (a_n). The power series one tends to consider are convergent for $|q|<1$. If $\sum_{n=0}^{\infty} a_n q^n$ is convergent for $-1 < q < 1$ then the *function* $f:(-1,1) \to \mathbf{R}$, $f(q) = \sum_{n=0}^{\infty} a_n q^n$, determines the sequence (a_n)

Note that we have the following identity of formal power series :

$$
\sum_{n=0}^{\infty} p(n) q^n = \prod_{n=1}^{\infty} (1 + q^n + q^{2n} + q^{3n} + \ldots)
$$

Indeed, the coefficient on the right-hand side is precisely the number of ways we can select non-negative integers n_i such that $\sum_{i=1}^{n} (n_i \cdot i) = n$, and these selections correspond to partitions of n. Every factor on the right is convergent for $|q|<1$ and so is the entire product. Hence

$$
\sum_{n=0}^{\infty} p(n) q^n = \prod_{n=1}^{\infty} (1 - q^n)^{-1}
$$

for $|q|<1$. As all similar identities are claimed to hold for the formal power series and for $|q|<1$, the condition $|q|<1$ is usually omitted, although it is assumed in all relations.

In addition to the partition function p(n) itself, one studies *restricted partition functions*. For a set U of partitions and $n \in \mathbf{N}$, let U(n) be the set of partitions of n that belong to U and set $p(U, n) = |U(n)|$. We shall consider the *restricted partition function* p(U, n) for a number of sets U. For example, O is the set of odd partitions, D is the set of partitions into distinct parts, P_o is the set of partitions into an odd number of parts, P_e is the set of partitions into an even number of parts, D_o is the set of partitions into an odd number of distinct parts, and D_e is the set of partitions into an even number of distinct parts. Thus $\lambda = (\lambda_i)_1^r$ belongs to O if each λ_i is odd and it belongs to D if $\lambda_1 > \lambda_2 > \ldots > \lambda_r$.

It is often helpful to represent partitions by their *diagrams* or *graphs* (see Figure 1). These diagrams, called *Ferrars* graphs or diagrams, were invented by N.M. Ferrars

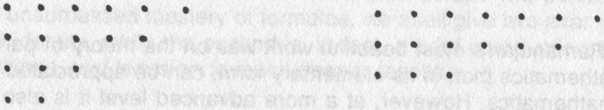


Figure 1. The Ferrars diagrams of (6,5,3,2) and (7,6,3)

and first appeared in print in a paper by J. J. Sylvester in 1853. It is often convenient to identify a partition with its Ferrars diagram.

The great Swiss mathematician Leonhard Euler started the study of partitions and the basic results were proved by him around 1740 and by Carl Gustav Jacob Jacobi about a hundred years later. Many of these results are fascinating and surprising identities of the kind p(U, n) = p(U′, n) for all n, where U and U′ are certain sets of partitions.

There are two natural ways open to us in proving a partition function identity: we may use function theory, relying on the generating function of the partitions, or we may use combinatorics, relying on the Ferrars diagrams of the partitions. We shall illustrate both approaches in proving some of the classical identities.

Theorem 1. The number of partitions of n into exactly m parts is equal to the number of partitions of n in which the largest part is m. The number of partitions of n into at most m parts is equal to the number of partitions of n into parts not greater than m.

Proof. Given a partition $\lambda = (\lambda_i)_1^r$ of n, let $S(\lambda)$ be the partition of n whose Ferrars diagram is obtained from the Ferrars diagram of λ by reflecting it in the main diagonal, as illustrated in Figure 2.

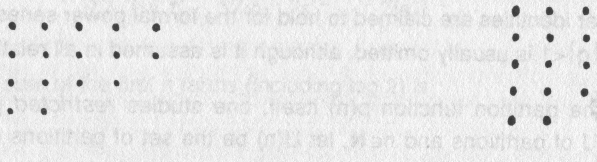


Figure 2. The Ferrars diagrams of $\lambda = (5,5,4,2)$ and $S(\lambda) = (4,4,3,3,2)$

It is immediate that S sets up a 1-1 correspondence between the appropriate sets : λ has exactly m parts if and only if m is the largest part of $S(\lambda)$. □

Now let us give a simple example of a partition identity whose proof is immediate if we use generating functions. This identity was proved by Euler in 1748.

Theorem 2. The number of partitions of n into odd parts is equal to the number of partitions of n into distinct parts.

Proof. What this result claims is that $p(O, n) = p(D, n)$ for every n. Note that

$$\sum_{n=0}^{\infty} p(O, n)\, q^n = \prod_{k=1}^{\infty} (1 + q^{2k-1} + q^{2(2k-1)} + \dots)$$

$$= \prod_{k=1}^{\infty} (1 - q^{2k-1})^{-1}$$

and

$$\sum_{n=0}^{\infty} p(D, n)\, q^n = \prod_{k=1}^{\infty} (1 + q^k) = \prod_{k=1}^{\infty} \frac{1 - q^{2k}}{1 - q^k}$$

$$= \frac{1 - q^2}{1 - q}\ \frac{1 - q^4}{1 - q^2}\ \frac{1 - q^6}{1 - q^3}\ \dots$$

$$= \prod_{k=1}^{\infty} (1 - q^{2k-1})^{-1}$$

where the last identity followed by cancelling the factors $1 - q^{2k}$. □

An identity fundamental in the theory of partitions is *Euler's pentagonal number theorem*. Just as the *triangular numbers* $k(k-1)/2$ count the number of elements in a triangular array (a Pascal triangle), the *pentagonal numbers* $k(3k-1)/2$ count the number of points in a pentagonal array (see Figure 3).

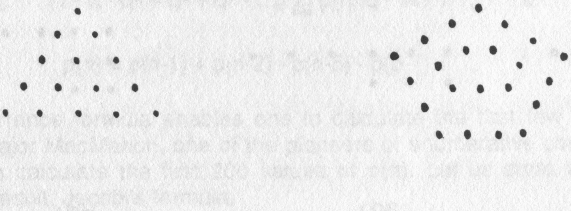


Figure 3. Representations of 6(6-1)/2 = 15 and 4(12-1)/2 = 22

As we shall see, the pentagonal number theorem makes it possible to give a useful recurrence relation for p(n).

Theorem 3.

$$\prod_{n=1}^{\infty} (1 - q^n) = 1 - q - q^2 + q^5 + q^7 - \dots = \sum_{-\infty}^{\infty} (-1)^k \; q^{k(3k+1)/2}$$

Proof. Let $\prod_{n=1}^{\infty} (1 - q^n) = \sum_{n=0}^{\infty} c_n q^n$. The theorem claims that $c_n = 0$ unless n is $k(3k-1)/2$ or $k(3k+1)/2$, in which case it is $(-1)^k$. What is c_n in terms of partitions? Clearly

$$c_n = p(D_e, n) - p(D_o, n)$$

because $\prod_{n=1}^{\infty} (1 + q^n) = \sum_{k=0}^{\infty} p(D, n)\, q^n$ is the generating function for distinct partitions, and a partition of n into m distinct parts contributes $(-1)^m$ to c_n. Thus we need an almost 1-1 correspondence between $D_e(n)$ and $D_o(n)$. We shall obtain such a correspondence by transforming the Ferrars diagrams.

Given $\lambda \in D$, the *slope* $s(\lambda)$ of the Ferrars diagram of λ is the maximal set of endpoints forming an internal sloping by 45° and containing the endpoint of the first line. The *base* $b(\lambda)$ is the set of points in the last line (see Figure 4).

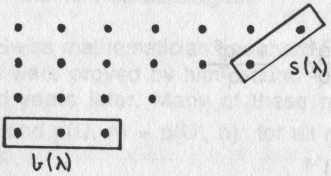


Figure 4. The Ferrars diagram of $\lambda = (7,6,4,3)$ with slope $s(\lambda)$, base $b(\lambda)$

Let us try two operations on (the Ferrars diagram of) a partition $\lambda \in D$. In operation β put the base next to the slope and in operation σ move the slope underneath the base. Call β *legal* if $\beta(\lambda) \in D$ and call σ *legal* if $\sigma(\lambda) \in D$ (see Figure 5).

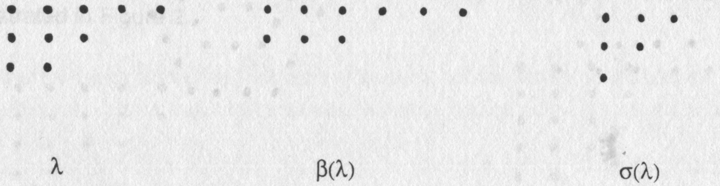


Figure 5. The operations β and σ on $\lambda = (5,3,2)$, with β illegal and σ legal

β is legal on λ if $|b(\lambda)| < |s(\lambda)|$ or if $|b(\lambda)| = |s(\lambda)|$ and $b(\lambda) \cap s(\lambda) = \varnothing$. Similarly, σ is legal on λ if $|s(\lambda)| < |b(\lambda)|-1$ or if $|s(\lambda)| = |b(\lambda)|-1$ and $b(\lambda) \cap s(\lambda) = \varnothing$.

Note that if β is legal on $\lambda \in D$ then σ is legal on $\beta(\lambda)$ and $\sigma(\beta(\lambda)) = \lambda$. Also, if σ is legal on $\lambda \in D$ then β is legal on $\sigma(\lambda)$ and $\beta(\sigma(\lambda)) = \lambda$. Furthermore, β and σ change the number of partitions by one so, if they are legal, they map an element of D_e into an element of D_o, and an element of D_o into an element of D_e.

When, then, is neither β nor σ legal on $\lambda \in D(n)$? If either $|b(\lambda)| = |s(\lambda)|$ and $b(\lambda) \cap s(\lambda) \neq \varnothing$ or $|s(\lambda)| = |b(\lambda)|-1$ and $b(\lambda) \cap s(\lambda) \neq \varnothing$. Setting $s = |s(\lambda)|$ we see that in the first case we must have $n = s + (s+1) + \ldots + (2s-1) = s\,(3s-1)\,/\,2$ and λ is unique, and in the second case $n = (s+1) + (s+2) + \ldots + 2s = s\,(3s+1)\,/\,2$ and λ is again unique (see Figure 6). Therefore if $n \neq s\,(3s \pm 1)\,/\,2$ then $c_n = p(D_e, n) - p(D_o, n) = 0$ since a combination of β and σ sets up a 1-1 correspondence between D_e and D_o. Furthermore, if $n = s\,(3s \pm 1)\,/\,2$ then $c_n = (-1)^s$ because the 'odd man out', the exceptional partition $\lambda \in D(n)$ on which neither β nor σ is legal, belongs to D_e if s is even and to D_o if s is odd. □

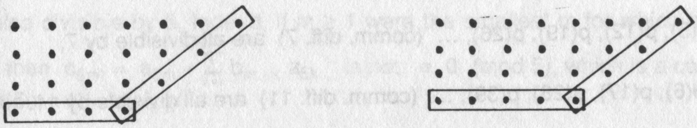


Figure 6. The exceptional Ferrars diagrams for n = 4(12-1)/2 =22 and n = 4(12+1)/2 = 26

Let us rewrite the pentagonal number theorem in a form that gives the promised recurrence formula for p(n).

Corollary 4. $(1 - q - q^2 + q^5 + q^7 - \ldots) \sum_{n=0}^{\infty} p(n)\, q^n = 1$, i.e

$$p(n) = p(n-1) + p(n-2) - p(n-5) - p(n-7) + \ldots$$ □

This recurrence formula enables one to calculate the first few values of p(n) rather rapidly. Major MacMahon, one of the pioneers of enumerative combinatorics, used this formula to calculate the first 200 values of p(n). Let us state without proof another classical result, Jacobi's formula.

Theorem 5.

$$\{\prod_{1}^{\infty} (1-q^n)\}^3 = 1 - 3q + 5q^3 - 7q^6 + \ldots$$

$$= {}^1/_2 \sum_{-\infty}^{\infty} (-1)^k \, (2k+1) \, q^{k(k+1)/2}$$

$$= \sum_{0}^{\infty} (-1)^k \, (2k+1) \, q^{k(k+1)/2}$$

☐

In 1918, from the sanatorium at Matlock, Ramanujan sent a letter to Hardy. "Dear Mr. Hardy, Will you please verify from the table of p(n) ($\sum p(n)\, x^n = (1-x)^{-1} (1-x^2)^{-1} \ldots$) the following?

(i) p(4), p(9), p(14), p(19), ... (comm. diff. 5) are all divisible by 5,

(ii) p(5), p(12), p(19), p(26), ... (comm. diff. 7) are all divisible by 7,

(iii) p(6), p(17), p(28), p(39), ... (comm. diff. 11) are all divisible by 11.

If these are true (especially (iii)) then ... "

The conjecture that followed, called, rather characteristically, a theorem by Ramanujan, did not get into the monumental partition paper Hardy and Ramanujan were writing at the time. However, Ramanujan did publish a proof of various results supporting this conjecture. To present a proof of the simplest of these results, we need a lemma. Let us write

$\sum a_n q^n \equiv \sum b_n q^n$ (mod m) if

$a_n \equiv b_n$ (mod m) for every n

Lemma 6. $\dfrac{1}{(1-q)^5} \equiv \dfrac{1}{1-q^5}$ (mod 5)

Proof. The result is immediate from the Binomial Theorem

Theorem 7. $p(5m+4) \equiv 0$ (mod 5)

Proof. Note that, by Jacobi's formula

$$q\{\prod_{1}^{\infty}(1-q^n)\}^4 = q\{\prod_{1}^{\infty}(1-q^n)\}^3 \{\prod_{1}^{\infty}(1-q^n)\}$$

$$= {}^1/_2\, q \sum_{-\infty}^{\infty} (-1)^k \, (2k+1) \, q^{k(k+1)/2} \sum_{-\infty}^{\infty} (-1)^l \, q^{l(3l+1)/2}$$

$$= {}^1/_2 \sum_{-\infty}^{\infty} \sum_{-\infty}^{\infty} (-1)^{k+l} \, (2k+1) \, q^{1 + k(k+1)/2 + l(3l+1)/2}$$

When is an exponent $\gamma(k, l) = 1 + k(k+1)/2 + l(3l+1)/2$ divisible by 5? Precisely when $8\gamma(k, l) - 10l^2 - 5 = (2k+1)^2 + 2(l+1)^2 \equiv 0 \pmod 5$. Now modulo 5 the first term is congruent to 0, 1, or -1, and the second term is congruent to 0, 2 or -2, so $\gamma(k, l) \equiv 0 \pmod 5$ if, and only if, $2k + 1 \equiv 0 \pmod 5$ and $l + 1 \equiv 0 \pmod 5$. In particular, the coefficient of q^{5m} in $q\{\prod_1^\infty (1-q^n)\}^4$ is a multiple of 5. Hence, by Lemma 6, the coefficient of q^{5m} in

$$q\{\prod_1^\infty (1-q^n)\}^4 \prod_1^\infty (1-q^{5n})(1-q^n)^{-5} = q\prod_1^\infty (1-q^{5n})(1-q^n)^{-1}$$

$$= q\{1+\sum_{n=1}^\infty b_n q^{5n}\}\{\prod (1-q^n)\}^{-1}$$

$$= \{\sum_1^\infty a_n q^n\}\{1+\sum_1^\infty b_n q^{5n}\}$$

is also divisible by 5. Note that $a_n = p(n-1)$ so $a_{5m+5} = p(5m+4)$. We claim that each a_{5m} is also divisible by 5. Indeed, if $m \geq 1$ were the smallest m for which not $a_{5m} \equiv 0 \pmod 5$ then $c_{5m} = a_{5m} + \sum_{k=1}^{m-1} b_{m-k}\, a_{5k}$ is not $\equiv 0 \pmod 5$, which is a contradiction. Hence $p(5m+4) = a_{5m+5}$. □

In fact, Ramanujan stated without proof two remarkable related identities, proved later by Darling and Mordell. The first of these was selected by Major MacMahon as the most beautiful identity discovered by Ramanujan. One need not be an expert in the field to appreciate its exotic beauty.

Theorem 8. $$\sum_{m=0}^\infty p(5m+4)\, q^m = 5\{\prod (1-q^{5n})\}^5 \{\prod (1-q^n)\}^{-6}$$

The full conjecture of Ramanujan turned out to be false and so it was modified. In 1967 Atkin proved this modified conjecture.

Theorem 9. If $d = 5^a 7^b 11^c$ and $24l \equiv 1 \pmod d$ then

$$p(l) \equiv 0 \pmod{5^a 7^{\lfloor (b+1)/2 \rfloor + 1} 11^c}. \quad \square$$

Note that in Theorem 7 we had $d = 5$ and $l = 5m + 4$.

It is a little strange that perhaps the most famous of all identities discovered by Ramanujan are the Rogers-Ramanujan identities. According to Hardy, these formulae are "certainly as remarkable as any which even Ramanujan ever wrote down". The discovery of these identities had been anticipated by a much less famous mathematician, L.J. Rogers. Ramanujan discovered these identities some time before 1913. He sent them to Hardy, who was very impressed by them, but could not prove them. Although Hardy asked several mathematicians about the formulae, nobody he asked could prove them or find a reference. While in Cambridge, Ramanujan attacked them again, but to no avail.

In the volume of his Harvard lectures (p.91), Hardy recalled what happened next. "The mystery was solved, trebly, in 1917. In that year Ramanujan, looking through old volumes of the Proceedings of the London Mathematical Society, came accidentally across Rogers' papers. I can remember very well his surprise, and the admiration which he expressed for Rogers' work. A correspondence followed in the course of which Rogers was led to a considerable simplification of his original proof. About the same time I. Schur, who was then cut off from England by the war, rediscovered the identities again."

What is very peculiar is that Rogers' reputation rests almost entirely on Ramanujan's rediscovery of the astonishing identities. Although Rogers proved them in 1894, he lost interest in the field and did not draw Ramanujan's attention to his work, perhaps because he did not even read Ramanujan's papers. Here are the two Rogers-Ramanujan identities.

Theorem 10.

$$\sum_{0}^{\infty} \frac{q^{n^2}}{(1-q)(1-q^2)\dots(1-q^n)} = \prod_{0}^{\infty} \frac{1}{(1-q^{5n+1})(1-q^{5n+4})}$$

$$\sum_{0}^{\infty} \frac{q^{n(n+1)}}{(1-q)(1-q^2)\dots(1-q^n)} = \prod_{0}^{\infty} \frac{1}{(1-q^{5n+2})(1-q^{5n+3})}$$

□

In fact, these identities can also be interpreted as rather natural relations concerning partitions. Let $C_{k,i}$ be the set of partitions $(\lambda_1,\dots,\lambda_r)$ such that $\lambda_j \not\equiv 0, \pm i$ (mod $2k+1$) for all j, and let $D_{k,i}$ be the set of partitions $(\lambda_1,\dots,\lambda_r)$ such that $\lambda_j - \lambda_{j+k-1} \geq 2$ and at most i-1 of the λ_j are 1. Then the first Rogers-Ramanujan identity states that

$$p(C_{2,2}, n) = p(D_{2,2}, n)$$

for all n. Indeed, the right-hand side is precisely

$$\sum p(C_{2,2}, n)\, q^n$$

What is the left-hand side? Take $\lambda = (\lambda_1,\dots,\lambda_m) \in D_{2,2}(n)$ so that $n = \sum_{1}^{m} \lambda_i$. Then $\lambda = (2m-1, 2m-3,\dots, 1) + \mu$, where μ is a partition of $n - m^2$ into at most m parts. Conversely, if μ is any partition of $n - m^2$ into at most m parts then $(2m-1, 2m-3,\dots, 1) + \mu \in D_{2,2}(n)$. Thus $D_{2,2}$ is enumerated by

$$\sum_{0}^{\infty} \frac{q^{m^2}}{(1-q)(1-q^2)\dots(1-q^m)}$$

In 1961 Gordon proved the following generalisation of the Rogers-Ramanujan identities.

Theorem 11. If $1 \le i \le k$ then $p(C_{k,i}, n) = p(D_{k,i}, n)$ for all n. □

Moreover, in 1981 Garsia and Milne proved a family of Rogers-Ramanujan type identities by using bijections, somewhat similar to the one used above in the proof of Euler's pentagonal theorem. Needless to say, the proof is much more complicated.

Let us turn to a rather different question concerning partitions. About how large is p(n) for large values of n? It is rather surprising that this had not been asked before 1917, when Hardy and Ramanujan published the amazing memoir on this topic we briefly referred to earlier.

The recurrence formula given in Corollary 4 enables us to calculate p(n) rather rapidly, provided we know the previous values, but it is of rather limited use when looking for an asymptotic formula. The function p(n) is rather unruly, it is certainly much more complicated than n!, so the readers who had to prove the rather simple Stirling approximation of n! in their first year at university, will appreciate the following result of Hardy and Ramanujan:

$$p(n) \sim (4n\sqrt{3})^{-1} \exp\{\pi (2n/3)^{1/2}\}$$

Impressive though this approximation is, it is hardly even the beginning. This formula was known to Ramanujan in India. But what about the error term? This is where the fun begins. Hardy and Ramanujan proved the following truly astonishing result.

Theorem 12. p(n) is the nearest integer to

$$(2\sqrt{2})^{-1} \sum_{q=1}^{\nu} \sqrt{q}\, A_q(n)\, \Psi_q(n) \qquad (*)$$

where

$$A_q(n) = \sum \omega_{p,q}\, e^{-2np\pi i/q}$$

the sum being over p's relatively prime to q and less than it, $\omega_{p,q}$ is a certain 24q-th root of unity,

$$\Psi_q(n) = {}^{d}/_{dx} \exp\{{}^{\pi}/_{q}\ ({}^{2}/_{3}\ (x - {}^{1}/_{24}))^{1/2}\}\ \big|_{x=n}$$

and ν can be taken to be of the order of $\sqrt{n}$.

In his review of *Ramanujan's Collected Papers*, Littlewood wrote about the extraordinary collaboration resulting in this amazing result.

> One of Ramanujan's Indian conjectures was that the first term of (*) was a very good approximation to p(n); this was established without great difficulty. At this stage the $x - {}^{1}/_{24}$ was represented by a plain x - the distinction is irrelevant. From this point the real attack begins. The next step in development, not a very great one, was to treat (*) as an 'asymptotic' series, of which a fixed number of terms (e.g. ν=4) were to be taken, the error being of the order of the

> next term. But from now to the end Ramanujan always insisted that much more was true than had been established: 'There must be a formula with error O(1).' This was his most important contribution; it was both absolutely essential and most extraordinary. A severe numerical test was now made, which elicited the astonishing facts about p(100) and p(200). Then ν was made a function of n; this *was* a very great step, and involved new and deep function-theory methods that Ramanujan obviously could not have discovered by himself. The complete theorem thus emerged. But the solution of the final difficulty was probably impossible without one more contribution from Ramanujan, this time a perfectly characteristic one. As if its analytic difficulties were not enough, the theorem was entrenched also behind almost impregnable defences of a purely formal kind. The form of the function $\Psi_q(n)$ is a kind of indivisible unit; among many asymptotically equivalent forms it is essential to select exactly the right one. Unless this is done at the outset, and the $-{}^1/_{24}$ (to say nothing of the ${}^d/_{dx}$) is an extraordinary stroke of formal genius, the complete result can never come into the picture at all. There is, indeed, a touch of real mystery. If only we *knew* there was a formula with error O(1), we might be forced, by slow stages, to the correct form of Ψ_q. But why was Ramanujan so certain there *was* one? Theoretical insight, to be the explanation, had to be of an order hardly to be credited. Yet it is hard to see what numerical instances could have been available to suggest so strong a result. And unless the form of Ψ_q was known already, *no* numerical evidence could suggest anything of the kind - there seems no escape, at least, from the conclusion that the discovery of the correct form has a single stroke of insight. We owe the theorem to a singularly happy collaboration of two men, of quite unlike gifts, in which each contributed the best, most characteristic, and most fortunate work that was in him. Ramanujan's genius did have this one opportunity worthy of it.

There are many other topics one should mention in connection with Ramanujan, including elliptic and modular functions, definite and indefinite integrals, continued fractions, elliptic theta functions, mock-theta functions and related q-series, the representation of numbers as sums of squares, and Ramanujan's τ-function. The last function is defined as follows:

$$\sum_{0}^{\infty} \tau(n)\, q^n = q \prod_{1}^{\infty} (1 - q^n)^{24}$$

Ramanujan conjectured that τ is multiplicative, i.e. $\tau(mn) = \tau(m)\tau(n)$ if m and n are relatively prime, but could not prove his conjecture. It was proved by L.J. Mordell, and the proof was not too difficult. Since then J.P. Serre and H.P.F. Swinnerton-Dyer used l-adic representations to place the whole area in a general setting, thereby proving wrong Hardy who, rather pessimistically, wrote that the area may be a backwater of mathematics. A much more important conjecture of Ramanujan, about the growth of $\tau(n)$, stating that $|\tau(n)| \leq 2\, p^{11/2}$ for every prime p, was generalised by Petersson and became known as the Ramanujan-Petersson conjecture. This was shown to be true by P. Deligne, when he proved the generalised Weil conjecture on algebraic varieties over finite fields.

During his stay in England, Ramanujan suffered much from his illness and from the climate. Nevertheless, he never regretted having made the journey and considered his five years in England as the greatest experience of his life. He published 21 papers in Europe, five of them in collaboration with Hardy.

A considerable body of Ramanujan's results is contained in three notebooks: one left with Hardy when he returned to India and two donated to the University of Madras upon his death. In their style these notebooks resemble the *Synopsis* of Carr read by Ramanujan with such enthusiasm. Professor B.C. Berndt has undertaken the task of editing these volumes: a difficult and arduous task indeed, since he intends to prove each of Ramanujan's theorems. Yet another notebook (in fact, about 100 loose sheets), named the 'lost notebook of Ramanujan' by Professor G.E. Andrews, contains Ramanujan's brilliant work on mock-theta functions. (The name 'lost notebook' is somewhat misleading since Andrews found it in the Wren Library of Trinity College, catalogued under 'Ramanujan'.) Ramanujan's sketchy results on mock-theta functions are among his most profound contributions to mathematics. The mathematical community owes a great debt to Professors Berndt and Andrews for their important work on these notebooks.

How great a mathematician was Ramanujan? Hardy thought that he could have been even greater: "The tragedy of Ramanujan was not that he died young, but that, during his five unfortunate years, his genius was misdirected, sidetracked, and to a certain extent distorted." This is, of course, perfectly possible. However, it is also possible that Ramanujan did find the field he was most suited for, in which he could make full use of his extraordinary formal powers. As Littlewood said, "every positive integer was one of Ramanujan's personal friends".

Ramanujan's amazing affinity to natural numbers is clearly illustrated by Hardy's well known and charming story about the taxicab. "I remember going to see him once when he was lying ill in Putney. I had ridden in taxicab No. 1729, and remarked that the number seemed a rather dull one, and that I hoped that it was not an unfavourable omen. 'No', he replied, 'it is a very interesting number; it is the smallest number expressible as a sum of two cubes in two different ways.' I asked him, naturally, whether he could tell me the solution of the corresponding problem for fourth powers; and he replied, after a moment's thought, that he knew no obvious example, and supposed that the first such number must be very large."

Hardy was rather fond of peculiar intellectual games like comparing mathematicians to cricketers and making up various elevens. In his 'Jerusalem Eleven' God was the opening batsman and Albert Einstein played at No. 3. Professor Paul Erdos, the prodigiously productive Hungarian mathematician who got to know Hardy and Littlewood in his early twenties, recalls Hardy's playful ratings of mathematicians on the basis of talent, the ratings going from 0 to 100. Hardy gave himself 25, Littlewood 30, Hilbert 80, and Ramanujan 100.

In pure, raw mathematical talent Ramanujan is unsurpassed.

* * *

Readers wishing to learn more about Ramanujan's mathematics are advised to start by consulting the following three volumes:

Hardy, G.H., Ramanujan, Twelve Lectures. Suggested by his Life and Work, University Press, Cambridge, 1940, 236 pp.

Andrews, G.E., The Theory of Partitions, Encyclopedia of Mathematics and its Applications, vol. 2, Addison-Wesley, Reading, Mass., 1976, xiv and 255 pp.

Berndt, B.C., Ramanujan's Notebooks, Part I, Springer-Verlag, New York, 1985, 357 pp.

ANSWERS TO PROBLEMS DRIVE

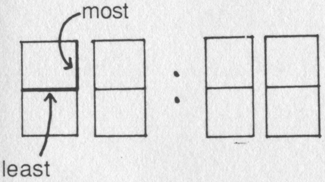
1. $1/\sqrt{3} + \sqrt{(2/7)}$

2. $2\sqrt{6}$ (cyclic quadrilateral)

3. Arthur and Edward are guilty.

4. AD = 5

5.

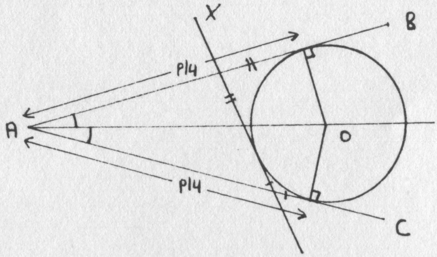


6. (i) 196 (twice sum of previous two terms)
   (ii) 1435 (2n! - n + 1)
   (iii) 20 (numbers beginning with the letter "t")
   (iv) 23 (primes in base 5)

7. Arc of circle centre A - length $\sqrt{(\pi/4\sqrt{3})}$

8. Least : tetrahedron, 700 miles
   Most : dodecahedron, 3900 miles

9.



10.

Fermat 1601-1665	Newton 1643-1727	Euler 1707-1783	Lagrange 1736-1813	Gauss 1777-1855	Galois 1811-1832
Riemann 1826-1866	Cantor 1845-1918	Hilbert 1862-1943			

11. $\pi/2$

12. 3975

The problems drive was won by Alan Stacey and Marcus Moore, who will set next year's questions.